2-1-2016

Nano-scale TG-FinFET: Simulation and Analysis

Ahmed Taha Elthakeb Naguib Youssef

The American University in Cairo

School of Science and Engineering (SSE)

Nano-scale TG-FinFET: Simulation and Analysis

A Thesis Submitted to

Electronics and Communications Engineering Department

in partial fulfillment of the requirements for

the degree of Master of Science

By Ahmed Taha El-Thakeb Naguib Youssef

Under the supervision of:

Prof. Yehea Ismail

Prof. Hamdy Abd Elhamid

July/2015

Cairo, Egypt

The American University in Cairo

School of Science and Engineering (SSE)

Nano-scale TG-FINFET: SIMULATION AND ANALYSIS

A Thesis Submitted by

Ahmed Taha El-Thakeb Naguib Youssef

To the Electronics and Communications Engineering Department

July/2015

In partial fulfillment of the requirements for

The degree of Master of Science

has been approved by

Thesis Committee Supervisor _____
Affiliation _____

Thesis Committee Reader_____
Affiliation _____

Thesis Committee Reader_____
Affiliation_____

_____ _____ _____ _____

Dept. Chair /Director Date Dean Date

DEDICATION

To my beloved family …

My father,

My mother,

and my brother.

ACKNOWLEDGMENT

Praise be to Allah, Lord of the Worlds

This work has been accomplished thanks to many persons. First, I would like to thank my supervisors Prof. Yehea Ismail and Prof. Hamdy Abd Elhamid for giving me the opportunity to carry out this work under their supervision.

I would like to thank Prof. Yehea Ismail for his trust and giving me the chance to join his center of nano-electronics and devices (CND) among a talented group of students and faculty. I appreciate his tremendous efforts to afford the most advanced research facilities including most recent T/CAD tools, PDKs and access to material experimental labs that greatly impacted the quality of this research. In addition to his insightful lectures and discussions that inspired me a lot in my research.

I would like to show my sincerest appreciation and gratitude to Prof. Hamdy Abd Elhamid. I would like to thank him for his sincere advising during which I have learnt a lot on both the technical and personal levels, for his fruitful discussions and continuous support and encouragement over two years of work. I also would like to thank him for giving me the opportunity to join the center of microelectronics, ICTEAM, UCL, Belgium, over the summer period which influenced both my research and life experiences.

I am greatly indebted to Prof. Hassan Mostafa for giving me the wonderful opportunity to investigate the FinFET on the circuit level. I would like to thank him for his precious time and valuable discussions and introducing me to the SRAM fundamentals.

I also would like to thank Prof. David Bol, UCL for his great advising and cooperation and nice hosting for me within his lab that helped me to make full use of my short stay. Also for providing me an access to the 28nm FDSOI PDK and carrying out very interesting study. I am also very grateful to Prof. Denis Flandre for his welcoming and insightful discussion that inspired me and was indispensable for our work.

I am obliged to my friend Taher Essam for his continuous assistance in installing the TCAD tools and his great help in the IT issues that saved a lot of time for me.

Finally, I would like to express my gratitude to my mother for her support and keeping my morals high in moments where desperation seemed to be the only choice. Also to all my friends in CND for their support and encouragement.

ABSTRACT

OF THE THESIS OF

Ahmed Taha El-Thakeb Naguib Youssef for Master of Science

Major: Electronics and Communications Engineering

The American University in Cairo

Title: Nano-scale TG-FinFET: Simulation and Analysis

Supervisor: Prof. Yehea Ismail Co-Supervisor: Prof. Hamdy Abd Elhamid

Transistor has been designed and fabricated in the same way since its invention more than four decades ago enabling exponential shrinking in the channel length. However, hitting fundamental limits imposed the need for introducing disruptive technology to take over. FinFET "3-D transistor" has been emerged as the first successor to MOSFET to continue the technology scaling roadmap.

In this thesis, scaling of nano-meter FinFET has been investigated on both the device and circuit levels. The studies, primarily, consider FinFET in its tri-gate (TG) structure.

On the device level, first, the main TCAD models used in simulating electron transport are benchmarked against the most accurate results on the semi-classical level using Monte Carlo techniques. Different models and modifications are investigated in a trial to extended one of the conventional models to the nano-scale simulations. Second, a numerical study for scaling TG-FinFET according to the most recent International Technology Roadmap of Semiconductors is carried out by means of quantum corrected 3-D Monte Carlo simulations in the ballistic and quasi-ballistic regimes, to assess its ultimate performance and scaling behavior for the next generations. Ballisticity ratio (BR) is extracted and discussed over different channel lengths. The electron velocity along the channel is analyzed showing the physical significance of the off-equilibrium transport with scaling the channel length.

On the circuit level, first, the impact of FinFET scaling on basic circuit blocks is investigated based on the PTM models. 256-bit (6T) SRAM is evaluated for channel lengths of 20nm down to 7nm showing the scaling trends of basic performance metrics. In addition, the impact of V_T variations on the delay, power, and stability is reported considering die-to-die variations. Second, we move to another peer-technology which is 28nm FD-SOI as a comparative study, keeping the SRAM cell as the test block, more advanced study is carried out considering the cell's stability and the evolution from dynamic to static metrics.

List of publications

- **Ahmed T. El-Thakeb**, Thomas Haine, Denis Flandre, Yehea Ismail, Hamdy Abd El Hamid, David Bol, "Analysis and Optimization for Dynamic Read Stability in 28nm SRAM Bitcells" in *IEEE ISCAS*, P. 1414-1417, May 2015.

- **Ahmed T. El-Thakeb,** Hamdy Abd El-Hamid, Yehea Ismail, "Scaling of TG-FinFETs: 3-D Monte Carlo Simulations in the Ballistic and Quasi-Ballistic Regimes," *IEEE Trans. Electron Devices*, vol. 62, no. 06, p.1796-1802, April 2015.

- **Ahmed T. El-Thakeb,** Hassan Mostafa, Hamdy Abd El-Hamid, Yehea Ismail, **"**Performance Evaluation of FinFET-Based SRAM Cell with Technology Scaling Under Statistical VT Variability,**"** i*n Proc. 26th Int. Conf. on Microelectronics (ICM)*, pp. 88–91, Dec. 2014.

In The Name of Allah, the Most Beneficent, the Most Merciful

Table of Contents

List of Figures

List of Tables

List of Abbreviations

MOSFET	Metal Oxide Semiconductor Field Effect Transistor
FinFET	Fin Field Effect Transistor
TG	Tri-gate
MG	Multi-gate
ITRS	International Technology Roadmap for Semiconductors.
DD	Drift-diffusion
HD	Hydro-dynamic
TD	Thermo-dynamic
MC	Monte Carlo
VT	Threshold voltage
SRAM	Static Random Access Memory
BTE	Boltzmann Transport Equation
BSIM-CMG	Standard compact models for multi-gate structures
PTM	Predictive technology models

1. INTRODUCTION

Transistors forming microprocessors, memory chips and telecommunications microcircuits have been designed in the same way since its invention at late 1950-1960's – what is called conventionally MOSFETs Figure 1-1. Basically, we may define transistors based on three design characteristics or metrics: a) the core design ***material*** involved in the fabrication process, b) the geometrical ***structure***, and c) the physical ***theory of operation*** describing its switching mechanisms between ON/OFF states, charge transport, and device electrostatics. Therefore, nano-electronics research centers and giant industry companies need to identify new materials, structures, and/or novel working principle in order to move forward.

Regarding the material, "Entire eras are named after materials — the Stone Age, the Iron Age and now we have the silicon age" said Shoucheng Zhang, a Stanford University physicist. The core material for fabricating mainstream transistors is Silicon so far, which has the ability to behave as both electrical conductor and insulator. However with the continuous miniaturization and looking forward to below 7nm channel lengths, moving to non-silicon CMOS may be prominent in the immediate future. There is a great interest in III-V high mobility materials for increased performance and higher switching speeds. In addition, a lot of efforts to integrate III-V materials with Silicon aiming to continue the scaling beyond the Silicon's capabilities alone. IMEC has already demonstrated world's first III-V FinFET devices monolithically with traditional Silicon substrate in late 2013 [*1*]. Also moving to the Carbon era is very likely with the great advances in Graphene materials [*2*].

Regarding the working principle, transistors are built on intrinsic substrate with two highly oppositely doped sides that are the source and drain. A channel in between connects the highly doped sides with a wide gate on top which controls the device

operation. Applying the right gate voltage, an inversion layer is formed in between that creates a conductive pathway that allows current to flow from source to drain.

Under this scenario the transistor is switched to be in the ON-state, while without forming such inversion layer, no current can flow hence the device is called to be in the OFF-state. In the OFF-state, what blocks the electrons flow is energy barrier across the channel, **Figure 1-1** and transistors are all about modulating these energy barriers through the gate and drain voltages. One of the fundamental problems impeding MOSFET scaling is the short channel effects (SCEs), where the energy barrier gets modulated not only by the gate but from the drain side as well, which compromise the idea working principle.

Extensive efforts are going on everywhere to identify new theories the switching mechanisms to take over [*3*]. Graphene bi-layers, which are simply two sheets of graphene in close proximity, are predicted to have special transport characteristics [*4*]. Another direction of interest is about spintronics and what is called spin-FET that basically makes use of the electron's spin to represent and process information [*5*].

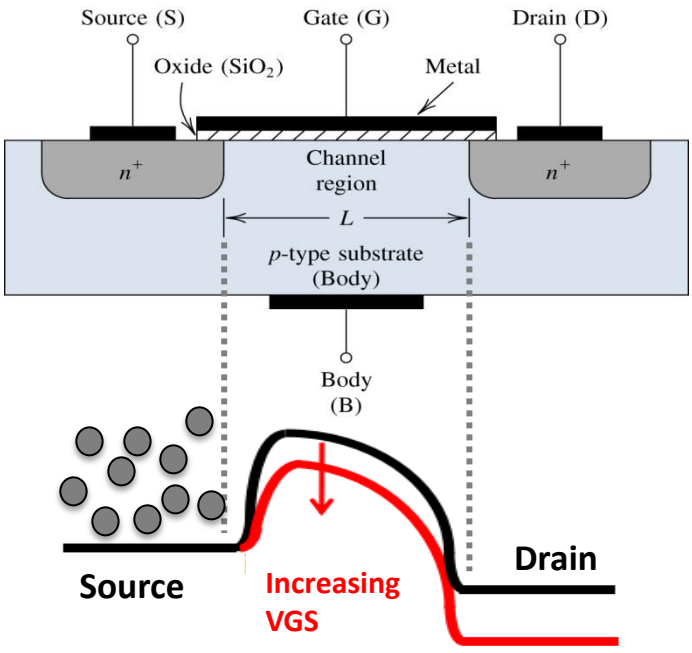


Figure 1-1: Illustratiing the MOSFET structure and its thwory of operation

1.1. Short channel effects (SCEs)

The most fundamental impediment associated with scaling down the channel length is what is called Short-channel effects (SCEs). SCEs are as a result of getting the source and drain closer to each other which result in undesired sharing of the electrical charges over the channel between the gate from one hand, and the source (S) and the drain (D) from the other side. The S and D junctions create a depletion region into the channel from each side, which effectively shorten the actual channel length under the gate control. As the drain voltage increases, more electric fields lines penetrates into the channel region and compromise the full control of the gate over the channel. The effect gets amplified as the distance between the S and the D gets shorter. As a resultant of losing the full gate control over the channel, two serious phenomena are observed that undermine the overall device performance, **Figure 1-2**:

a) Drain-induced barrier lowering (DIBL); which causes the threshold voltage to decrease with increasing the drain voltage.
b) Degradation in the sub-threshold slope (SS).

Both of them contribute to increase the overall leakage current of the transistors forming a serious challenge for further technology scaling.

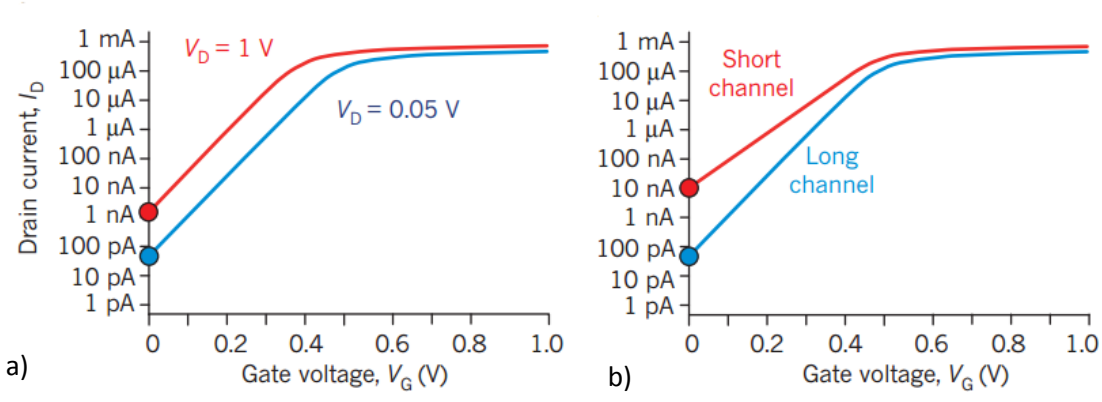


Figure 1-2: Illustration of short-channel effects. a) DIBL effect, b) Sub-threshold swing degradation

1.2. Tri-gate “FinFET” structure

In the conventional MOSFET structure, the source, the drain, and the channel connecting them (with the gate on top) lie flat in the same plane. In such configuration the device is called planar and can essentially be treated as a 2-D device especially on the simulation level. In this case, the electrostatic control is achieved through a capacitive coupling between the top gate and the channel region though the gate oxide layer. SCEs can be reduced by improving the gate control over the channel which can be achieved through two approaches. First, increasing the gate control by enhancing the capacitive coupling between the gate and the channel through the reduction of the gate oxide thickness or using high-k oxides. Second, reducing the impact of the drain by decreasing the depth of the source and drain regions with scaling the channel length.

On the other side, device’s electrostatics can also be improved by modifying the shape of the device. For long channel MOSFETs, the device’s electrostatics are considered as a one dimensional problem and the gradual channel approximation was commonly employed in the old days in solving 1-D Poisson equation (that govern the relationship between the electric fields and the charges in the vertical direction. Having SCEs when electric fields from both sides (S/D) penetrate laterally (in the horizontal direction) into the channel, extends the problem into a 2-D problem.

Multi-gate (MG) structures (also called FinFETs) make use of the third dimension to mitigate SCEs by increasing the gate control over the channel region. MG structures come in different flavors as shown in Figure **1-3**, in the next section we will explain the effect of increasing the gate area on the natural screening length. Basically the name comes from the shape of the gate and how many sides it wraps around [*6*]. Also the substrate can be SOI or Bulk.

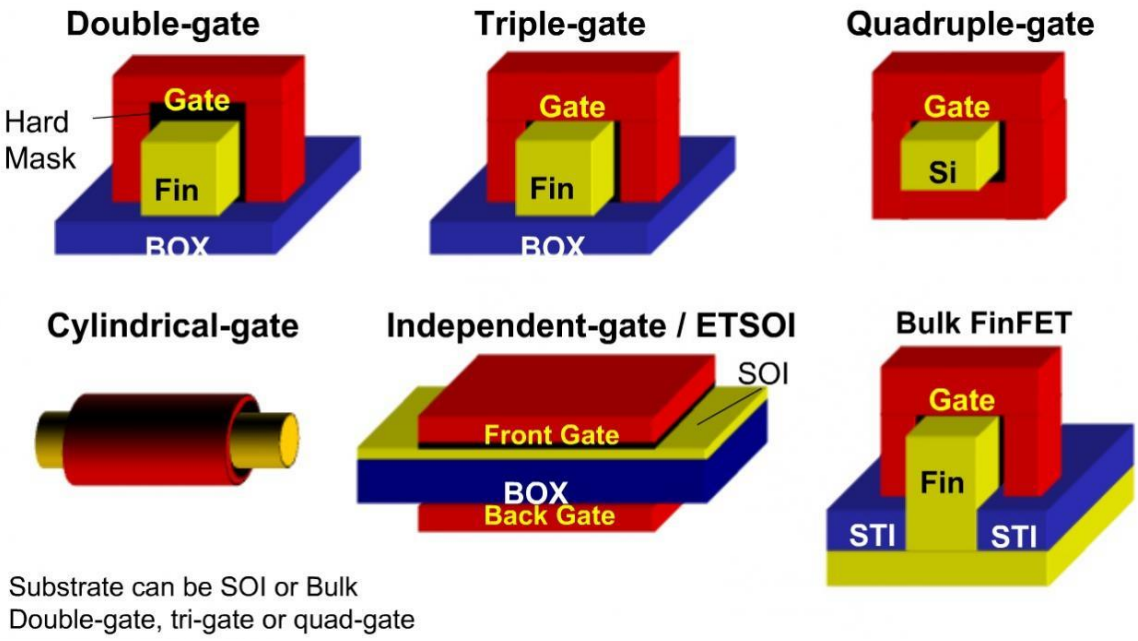


Figure 1-3: Different flavors of MG structures

1.3. Reduction of short-channel effects

SCEs, DIBL and SS degradation, are mainly a result of the penetration of the electric field lines from the drain end into the channel hence competing the gate in modulating the energy barrier and consequently becomes more difficult to turn the device off by reducing the threshold voltage. Maxwell's equation describes the distribution of the electric potential along the channel [*7*]:

$$\nabla . D = \rho \tag{1-1}$$

where $D = \varepsilon E$ is the electrical displacement field, ε is the permittivity of the material, and E is the electric field and ρ is the local density of electrical charge. In 3-D, the electric field components are shown in Figure **1-4**, and Poisson equation is written as:

$$\frac{dE_x}{dx} + \frac{dE_y}{dy} + \frac{dE_z}{dz} = -\frac{\rho}{\varepsilon} = \text{consant value}$$

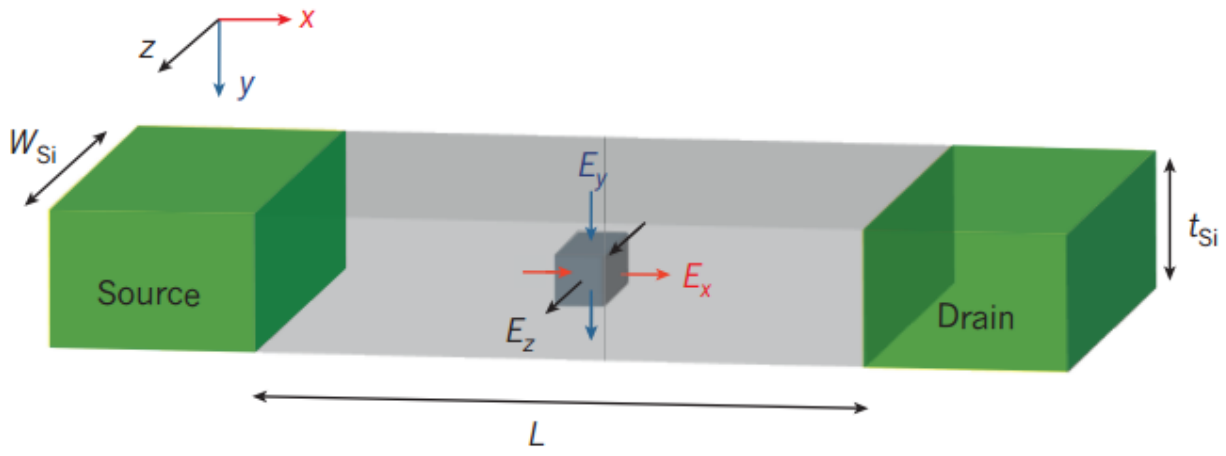


Figure 1-4: Different Electric Field components on elemental volume inside the channel

The superiority of the multi-gate structures over the planar is that the gate control (in the MG case) is exerted in the y and z directions and competes with the undesired variation in the electric field in the x direction coming from the source and the drain.

The sum of three components in x, y, and z of the Poisson equation is a constant, therefore any increase in the control by the top and bottom gates (dE_y/dy), or by the side gates (dE_z/dz) will counteract the SCEs by reducing the penetration of the electric field component coming from the S and D (dE_x/dx).

Using simplifying assumptions and few approximations, it is possible to deduce from solving the Poisson equation a parameter called the geometric screening length (or the natural length) which represents the extension of the electric field lines from the source and the drain into the channel region [*6*]. For example, to quantify such parameter, it is possible to get a device free of SCEs if its gate length is times larger than the natural length (L>6 λ).

For a device with a square cross section having width of W and thickness of T, the natural length is given by:

$$\lambda_1 = \sqrt{\frac{\varepsilon_{si}}{\varepsilon_{ox}}} t_{ox} t_{si} \qquad \text{For single gate MOSFET} \qquad (1\text{-}2)$$

$$\lambda_2 = \sqrt{\frac{\varepsilon_{si}}{2\varepsilon_{ox}}} t_{ox} t_{si} \qquad \text{For double gate MOSFET} \qquad (1\text{-}3)$$

$$\lambda_4 = \sqrt{\frac{\varepsilon_{si}}{4\varepsilon_{ox}}} t_{ox} t_{si} \qquad \text{For quadruple gate MOSFET (GAA)} \qquad (1\text{-}4)$$

where ε_{ox} is the electrical permittivity of the gate oxide, ε_{Si} is the electrical permittivity of the silicon, t_{ox} is the gate oxide thickness, and t_{Si} is the silicon film thickness. These expressions indicate that the SCEs can be minimized by decreasing the gate oxide thickness, by decreasing the silicon film thickness, and by increasing the dielectric constant of the gate oxide material.

Looking at the natural lengths of different MG structures, equations (1-3), interesting concept can be defined as the effective gate number (N), and a generalized natural length expression can be written in terms of N as follows:

$$\lambda_N = \sqrt{\frac{\varepsilon_{si}}{N\varepsilon_{ox}}} t_{ox} t_{si} \qquad (1\text{-}5)$$

This expression clearly shows the benefits of the MG structures in improving the device electrostatics by reducing the SCEs.

1.4. Overview of the Thesis

As it is clear from the above introduction, the focus of this research is multi-gate structures and specifically Tri-gate (TG) FinFET. The thesis is composed of two main parts: A) Device Level [Ch.2, 3], B) Circuit Level [Ch.4].

On the device level; the main focus is about the electron transport in nano-scale TG-FinFET and is divided into two parts [Ch.2, Ch3].

In Chapter 2, the basic electron transport models are discussed starting from the most classical drift-diffusion model to most sophisticated Monte Carlo techniques. A case study of double-gate FinFET structure is used to show the failure of the conventional DD model in simulating nano-scale channels. Then, the main transport models are benchmarked against the most accurate results on the semi-classical level from Monte Carlo techniques. Different models and modifications are investigated in a trial to extended one of the conventional models to the nano-scale simulations, since they are relatively simple and computationally efficient compared to the Monte Carlo ones.

In Chapter 3, using the conclusions from the previous chapter, a numerical study for scaling nano-scale TG-FinFET according to the most recent International Technology Roadmap of Semiconductors (ITRS 2013) is carried out by means of 3-D Monte Carlo simulations in the ballistic and quasi-ballistic regimes. Ballisticity ratio (BR) is extracted and discussed over different channel lengths. The electron velocity along the channel is analyzed showing the impact of spatial portions of the channel on the transport behavior.

In Chapter 4, we start by benchmarking the basic performance of TG-FinFET SRAM cell (256-array) with technology scaling starting at 20nm and down to 7nm channel length. In this study, predictive technology models (PTM-models) are used as the model cards for the simulations with BSIM-CMG models (the standard compact models for MG-FETs developed by BSIM group).

Next, we move to another peer technology which is 28nm FD-SOI as the most advanced available commercial PDK, keeping the SRAM cell as the test block, but more advanced study is carried out about the cell stability and the evolution from the dynamic to the static metric.

Finally, in Chapter 5, we conclude the overall results of this research and propose future work.

2. BENCHMARKING SEMI-CLASSICAL ELECTRON TRANSPORT MODELS FOR NANO-SCALE FINFET IN TCAD

2.1. Introduction

Device models can be categorized under one of three different types of models: a) TCAD models, b) empirical models, and c) compact models. The first type, TCAD, are based on numerical solving techniques solve for the carrier transport and electrostatics of different devices in exact manner; however its computational burden increases especially after the technology advances brought new device architectures such as the FinFET which requires three-dimensional simulations making them impractical for fast circuit simulation, yet, they are of extreme importance for rigorous device physics analysis, so some techniques need to be implemented to turn them to be more efficient. On the other side, the second and third types of models have much less computational burden so they are more practical for fast circuit simulation, however, for the second type, as the dimensions shrink, the complexity of having novel geometries and new physics of carrier transport such as hot electrons phenomena, velocity overshoot, ballistic/quasi-ballistic transport, and quantum effects, impose an enormous number of empirical parameters to be used in the model that drives them far from physical and consequently reduce the amount of insights out of them. All these complexities underlying technology scaling imply the need for more understanding and analyses for the physics involving the transport at the nano-scale dimensions, which consequently can yield more physics-based compact models that can predict the device performance properly and are computationally efficient at the same time [*8*]. In the following section, we discuss the computational electronics which enables combining sophisticated numerical techniques along with physical models and incorporating them efficiently into TCAD tools for advanced simulations of semiconductor devices.

2.2. Computational electronics

Modeling and even simulation of nano-scale FinFET is a formidable challenge due to several factors. At such extremely scaled dimensions it becomes more and more complicated to understand the actual operation of such devices specifically from the electron transport point of view, since peculiar effects start to show up at these extremely scaled dimensions such as hot electrons, velocity overshoot, ballistic and quasi-ballistic transport [*9*], [*10*], [*11*]. Therefore careful treatment for the electronic transport must be considered. These facts imply that relying on fully experimental approach that encounters trial and error will be impossible in terms of both time consumption and cost.

Relying on the technology advancements enabled by the electronics so far, computers are considered cheaper and more practical resources to address the analysis and simulation of further technology nodes and practically become an indispensable tool for all device engineers.

Computational Electronics is devoted to state of the art numerical techniques and physical models used in the simulation of semiconductor devices from a semi-classical perspective and can be extended to include more advanced physics such as quantum transport which is the base of Technology Computer Aided Design (TCAD) tools. In fact its importance mainly stems from two points: a) offering the possibility to investigate physical phenomena that cannot be measured in real life experiments which offers much insight into the real theory of operation of the device under test, b) it enables examining novel devices or even hypothetical devices which have not been manufactured yet [*12*].

In addition, this kind of simulations can include process simulation that consider various device fabrication processes such as oxidation, etching, material deposition and growth, impurity diffusion, contact deposition. TCAD provides the basis for device modeling as the SPICE simulators provide the basis for circuit design.

The main design flow steps to achieve specific customer need are shown in Figure **2-1**. The basic components for general semiconductor device simulation are shown in Figure **2-2**.

The basic methodology can be described in terms of two coupled kernels that need to be solved self-consistently with each other, a) the transport equations that govern the flow of charge carriers, b) the electrostatics which describes the modulation of energy barriers and essentially drives the charge flow. Both of them are coupled to each other therefore they require simultaneous solution. Initially, with the beginning of the semiconductor industry, the electrical device characteristics were estimated using pure simple analytical models, for example, the gradual channel approximation for MOSFETs based on the drift-diffusion model which encountered several approximations to yield closed form expressions. The resulting formulas, however, were able to capture the basic device behavior and features [*13*]. Examples of such approximations are using simplified doping profiles and structure geometries. However, with the advancement in the semiconductor industry and the continuous shrinking of the channel length, these approximations can no longer be applied and start to lose its validity. Hence there was a need for more accurate models.

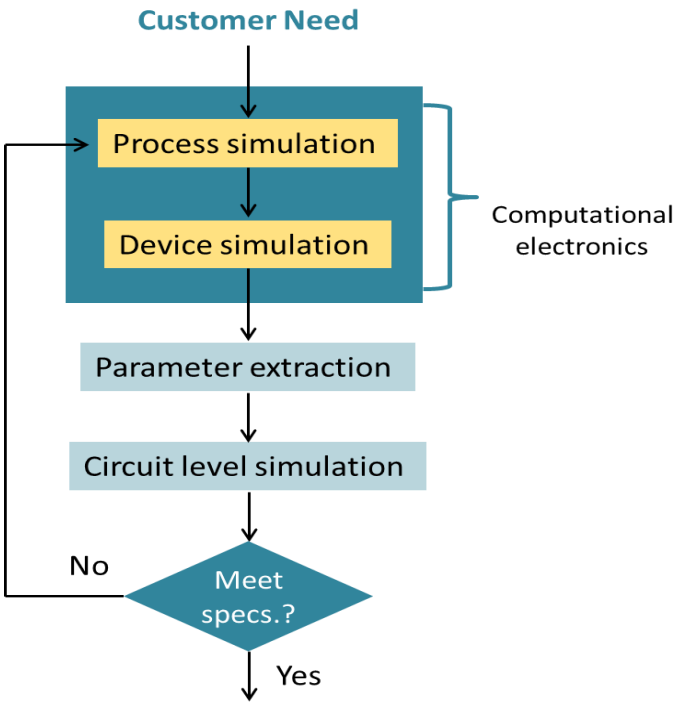


Figure 2-1: Design sequence to achieve desired customer need

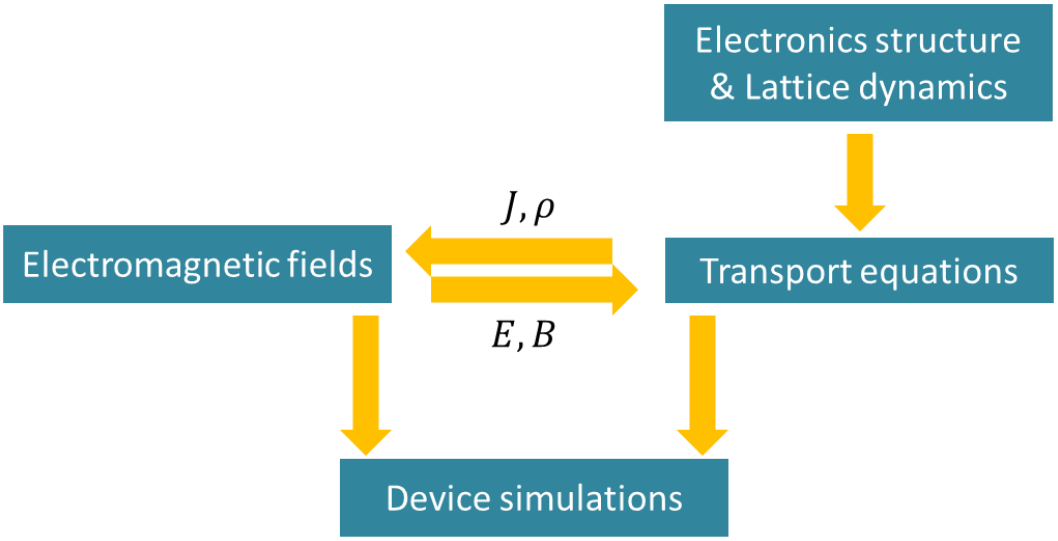


Figure 2-2: Sequence of main device simulation

Numerical simulation was the base for developing such advanced models by solving the carrier transport of semiconductor devices using discretization as was demonstrated by the work of scharfetter and Gummel [*14*] who proposed a robust discretization of the DD equations that are still in use till today. In the next section, we move to describe the main carrier transport models used in simulating transistors and their evolution over the past decades till today.

2.3. Electron Transport models

Modeling of carriers under equilibrium (rest state) conditions is necessary since it establishes the initial frame of reference. However, under equilibrium the net current flow is zero which is uninteresting for practical performance demonstration. Therefore, from a device performance point of view, when the semiconductor is excited, this gives rise to carrier action or a net carrier response and essentially current can flow. So the most interesting question, what controls the operation of the semiconductor devices, is how the charge carriers (electrons/holes) respond to applied, built–in, and or scattering potentials.

In fact, with shrinking the channel length, the clear understanding of carrier transport remains the most tedious issue for proper device modeling [*15*].

In semiconductor devices, there are two types of carrier motion as shown in Figure **2-3**, A) deterministic motion where electrons can be considered as a classical particles hence the Newton's law can be applied, and B) random motion; since after some time, the electron encounter a scattering event which essentially changes its direction and momentum, these random scatterings events follow Fermi's golden rule. As it can be noticed in the same figure, when l (denoting the effective channel length) is much longer than the mean free path (λ), the transport is mainly described as drift and diffusion components. As l scales down, the transport becomes more deterministic due to the reduction of the number of the scattering events that induce this randomness since the device becomes shorter. Therefore, the transport goes from drift-diffusion to quasi-ballistic (l ~ λ) and eventually to ballistic at (l < λ) as will be described in the subsequent sections.

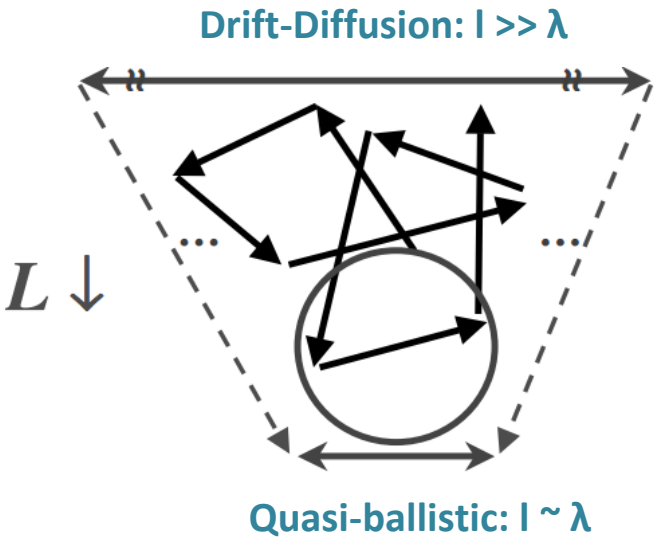


Figure 2-3: Illustration of carriers' motion inside a semiconductor. Each arrow represents a deterministic path until an abrupt change or scattering event happens so the carrier changes its momentum randomly and goes through another deterministic path represented by different arrow, and so on.

2.3.1. Drift-Diffusion (DD) Transport Model

The most popular transport model that has been used over long period and all device engineers rely on is called the drift-diffusion model (DD). DD represents one of the semi-classical approaches of treating carrier transport in semiconductors and is based on macroscopic theory in a sense that it considers the electrons as particles.

In the normal case, under equilibrium, electrons execute random thermal motion, where they move in a direction for a while until they encounter a scattering event which essentially changes their direction. Examples for such scattering events could be due lattice vibrations or impurity scatterings and many others. This scattering process might result-in a change in the momentum and/or the energy of the charge carrier.

Since this is under equilibrium, the net current flow is zero, however the electrons have thermal kinetic energy (KT) and average thermal velocity (v_{th}~$10^7\ cm/s$). Having particles exhibiting a random walk, statistical approaches are used to characterize their behavior such as Fermi and Boltzmann statistics.

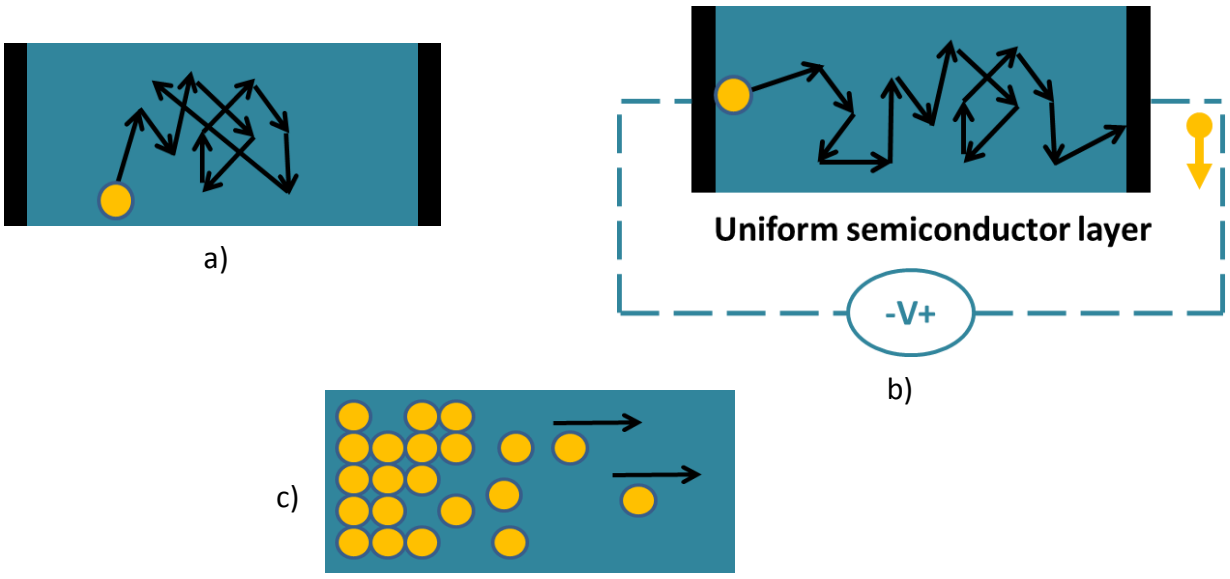


Figure 2-4: Drift Diffusion transport mechanisms: a) random walk under thermal equilibrium, b) Drift under applied electric field, c) Diffusion under concentration gradient.

DD model considers the carriers' motion consists of basically two components: i) Drift; where the current carriers drift under the influence of electric field, Figure 2-4, b).ii) Diffusion; where charge carriers diffuse down under concentration gradient, Figure 2-4 c). In general, we have both concentration gradient and electric fields. Hence DD model can be described by this general equation:

$$J_n = n_s q \mu_n E + q D_n \frac{dn_s}{dx} \quad (2\text{-}1)$$

$$\frac{D_n}{\mu_n} = \frac{KT}{q} \quad (2\text{-}2)$$

where J_n is the electron current density, q is the electron charge, n_s the electron charge density, μ_n is the electron mobility, E is the electric field, D_n is the electrons diffusion coefficient, $\frac{dn_s}{dx}$ is the electrons concentration gradient, T is the temperature, and K is the Boltzmann constant. DD model is based on the first moment of the BTE, and is strictly valid for low field near equilibrium conditions found in long channel transistors [*16*]. But, with scaling the channel length, the DD model starts to lose its validity since some of the assumptions of this macroscopic model that are implemented in the TCAD tools start to break down. The first is assuming collision dominated transport, and the second is neglecting the quantum effects and the degenerate carrier statistics [*17*].

In addition, it was found that he DD model underestimates the ballistic on-current [*18*] due to its incorrect limit on the carrier velocity and shows no velocity overshoot due to its local transport assumption.

2.3.2. Thermodynamic (TD) Transport Model

The thermodynamic transport model extends the drift-diffusion approach to account for electro-thermal effects, under the assumption that charge carriers are in thermal equilibrium with the lattice.

Therefore, the carrier temperatures and the lattice temperature are described by a single temperature. The thermodynamic model is required for simulations with high current

levels, where considerable self-heating effects might occur. Examples for such cases can include power devices and MOSFETs with high gate or drain voltages, and open bipolar transistors.

The reason behind this model is that high currents can produce Joule heat in the device's regions, which may raise the lattice temperature significantly.

Since many models used in the simulations, including the carrier mobility models, the SRH generation-recombination models, and the avalanche generation model, are functions of the lattice temperature, solving the heat flow equation (thereby obtaining the lattice temperature distribution) is necessary to improve the accuracy of the simulation under such conditions.

The thermodynamic model can be used independently or combined with other advanced transport models [*19*].

In practice it solves the lattice temperature (heat flow) equation in addition to Poisson equation and carrier continuity equation.

The thermodynamic model is defined by the basic set of differential equations after adding the temperature gradient [*19*]:

$$J_N = qn\mu_n E + qD_N \nabla n - qn\mu_n P_n \nabla \mathrm{T} \quad (2\text{-}3)$$

$$J_P = qp\mu_p E - qD_P \nabla p - qp\mu_p P_p \nabla \mathrm{T} \quad (2\text{-}4)$$

where P_p and P_n are the absolute thermoelectric powers, and $\nabla \mathrm{T}$ is the temperature gradient.

2.3.3. Hydrodynamic (HD) Transport Model

In relatively small channel lengths, the carriers move through the device with velocity larger than the saturation velocity which induce a non-stationary kind of transport and non-local effects where the mobility becomes field dependent.

In Si devices non-stationary transport occurs because of the different order of magnitude of the carrier momentum and energy relaxation times.

Hydrodynamic model was developed mainly to investigate such non-stationary and non-equilibrium electron transport in sub-micrometer channel transistors and semiconductor device [*20*].

The HD model gained its popularity in electron transport theory due to the physical features of this approach in addition to its practical attributes. In Hydrodynamic/Energy balance modeling the velocity overshoot effect is accounted for through the addition of:

- Energy conservation equation, in addition to:
- Particle Conservation (Continuity Equation)
- Momentum (mass) Conservation Equation

which is the superiority of the HD over the classical DD model. [*20*].

Another model called Energy transport model (ET) which is usually mentioned when discussing HD models. The basic difference between the HD and ET models is the neglect of the drift energy in the energy transport equation [*21*].
However, with the continuous scaling of channel length approaching the near ballistic regime, it was found that both ET/HD may substantially overestimate the on-current [*21*]. One justification for the failure of such macroscopic transport models in the near ballistic can be attributed to the assumption of their derivation. The kinetic energy of carriers is composed of two terms, the first is the thermal energy due scattering events, and the second is the drift energy associated with average motion of the carriers. In such models, it is common to neglect the second term. However, at the ballistic limit, there is no scattering to rise the temperature, hence the second term dominates the total kinetic energy. Neglecting the drift energy term in such models is most probably the cause of the un-physically high velocities observed in the HD/ET simulations [*21*].

2.4. Benchmarking semi-classical transport models in TCAD

2.4.1. Problem statement

As discussed above about the complexity associated with scaling down the channel length and the evolution of new carrier transport physics, conventional models can no longer be used to simulate such nano-scale devices. Therefore special care should be devoted to choosing the proper transport model of simulation. To give an idea about the importance of this point, a double gate (DG) structure, as shown in Figure **2-5** a), was simulated with two different channel lengths: a) L=50 nm, b) L=20 nm, representing long and short channels respectively. The doping profile is shown in Figure **2-5** b), and the main device parameters and dimensions are summarized in Table **2-1**. The simulation was done two times using different transport model in each run:

i) Conventional Drift-diffusion,
ii) Monte Carlo technique (will be discussed in more details in the following sections).

to assess validity of the used carrier transport model with scaling the channel length.

Figure 2-6 a), b) show the transfer characteristics of the simulated device in the saturation regime ($V_d = V_{DD}$), at each channel length for both DD and MC.

It is clear that for quite long channels (L=50 nm), as shown in Figure 2-6 a), the MC and DD models quite match each other and yields almost the same results while scaling the channel length down to 20 nm, Figure 2-6 b), the DD model clearly underestimates the on-current and a big mismatch is found with respect to the characteristic obtained using the MC model. These results are consistent with previous studies [*18*].

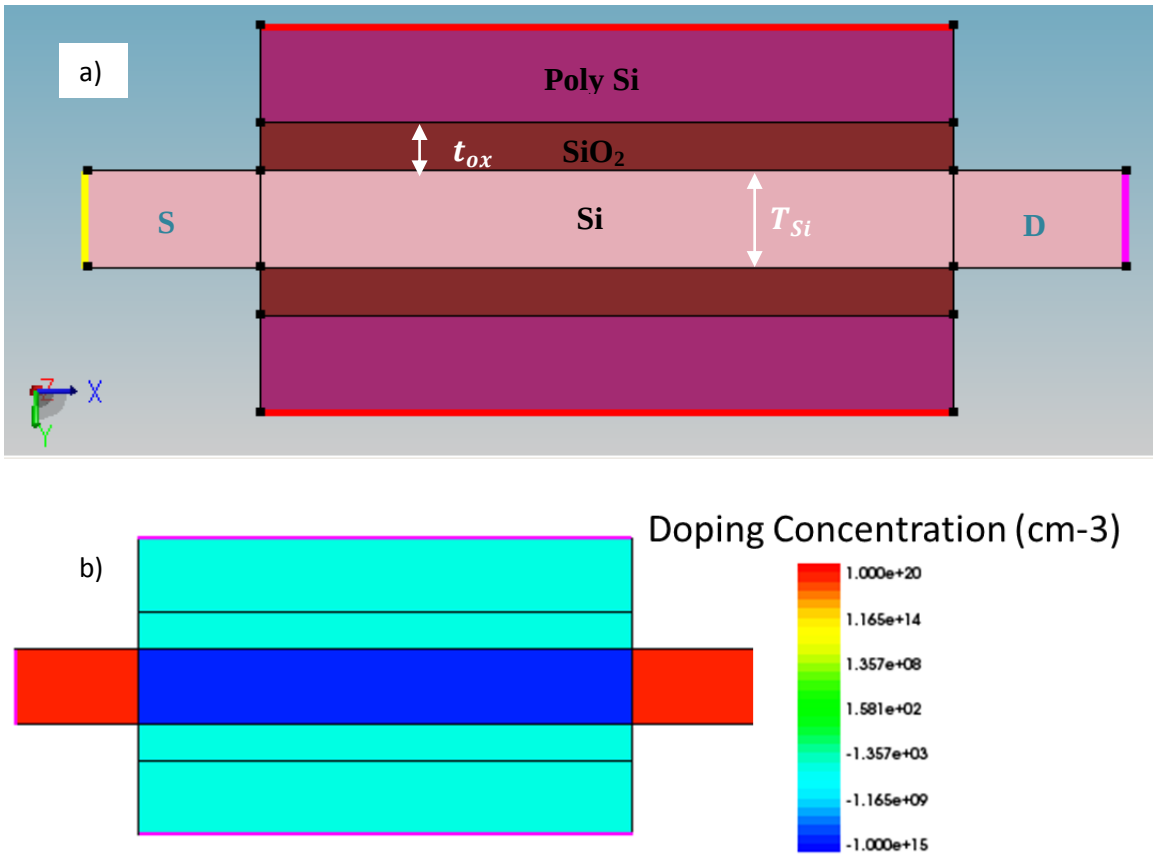


Figure 2-5: Simulated double gate (DG) structure, (a) Structure's geometry by Sentaurus structure editor, (b) Doping profile.

Table 2-1: Device parameters of the simulated structure

Parameter	Value
Channel length (L)	a) 50 nm b) 20 nm
Body Thickness (T)	3 nm
Oxide Thickness (t_{ox})	1.5 nm
Body Doping (NA)	-1e15
S/D Doping (ND)	1e20

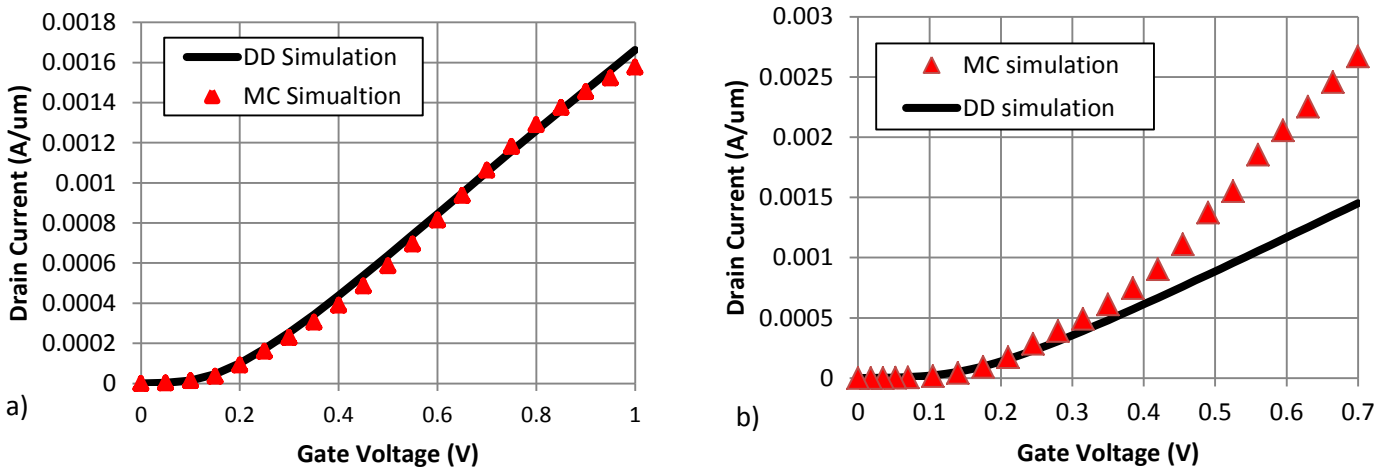


Figure 2-6: Transfer characteristics with Monte Carlo (MC), and classical drift-diffusion (DD): (a) long channel $\boldsymbol{L_{eff} = 50nm}$, (b) short channel, $\boldsymbol{L_{eff} = 20nm}$

2.4.2. Objective of the study

The purpose of this study is to investigate and assess the applicability of the common electron transport models used in commercial TCAD device simulator (Sentaurus) to describe the behavior of nano-scale channel lengths, where the quasi-ballistic regime is dominant [2 - Bude], with novel structural geometries such as triple-gate (TG) FinFETs at the dimensions projected by the international technology roadmap for semiconductors (ITRS).

2.4.3. Device Structure and Simulation Methodology

As a case study, TG FinFET structure is used according to the process in [*22*] for two channel lengths: a) L=17 nm, b) L=15nm. The main process formation flow is shown in Figure 2-7, and the doping profile is shown in Figure 2-8, and finally Figure **2-9**, shows the structure after meshing (representing the geometrical object as a set of finite elements for computational analysis). Different versions and modifications of conventional Drift-

Diffusion (DD) model are investigated. Classical Monte Carlo simulations were taken as a reference since they are considered the most accurate results on the semi-classical level.

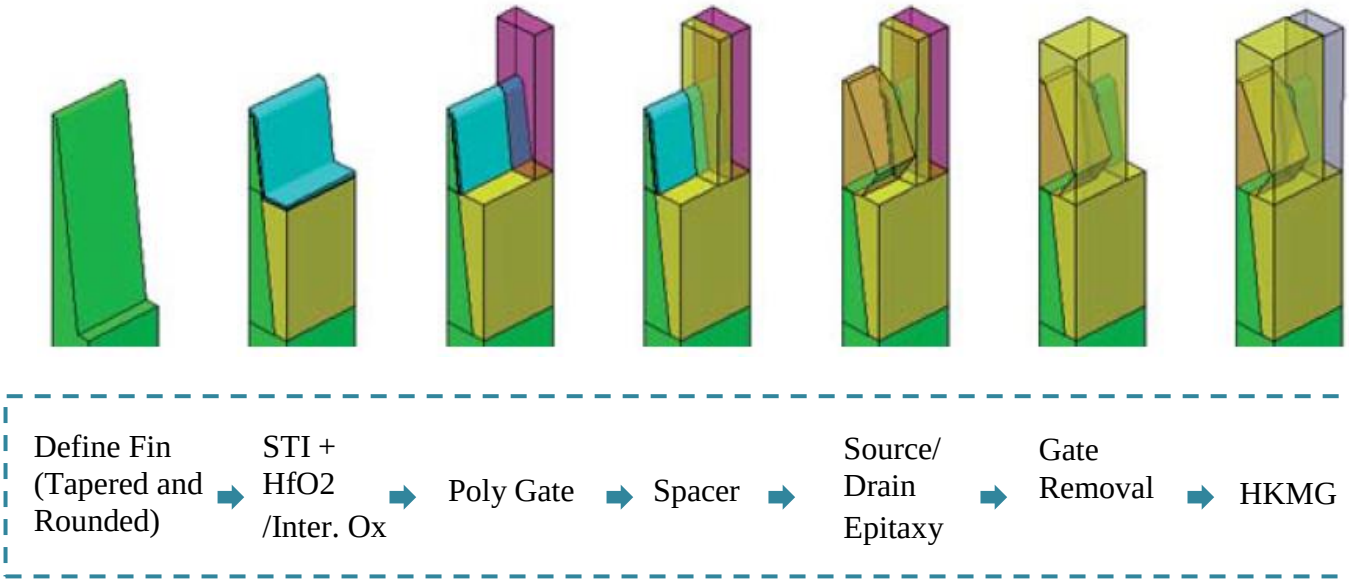


Figure 2-7: Process Formation Flow of simulated device [Sentaurus Template, ***[81]***]

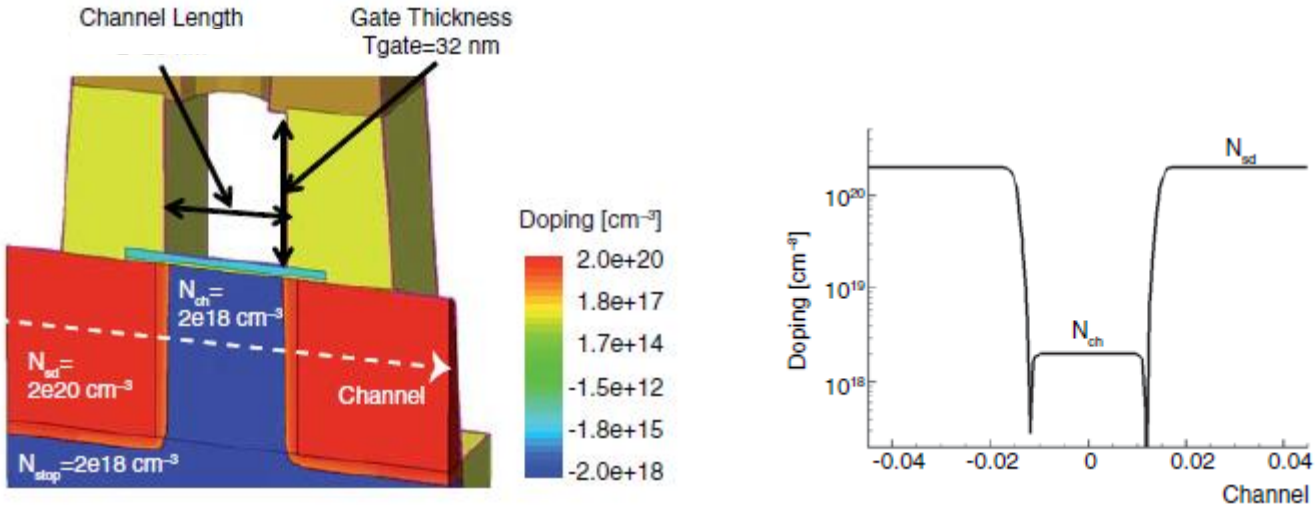


Figure 2-8: Doping Profile across the simulated structure, [Sentaurus Template, ***[81]***]

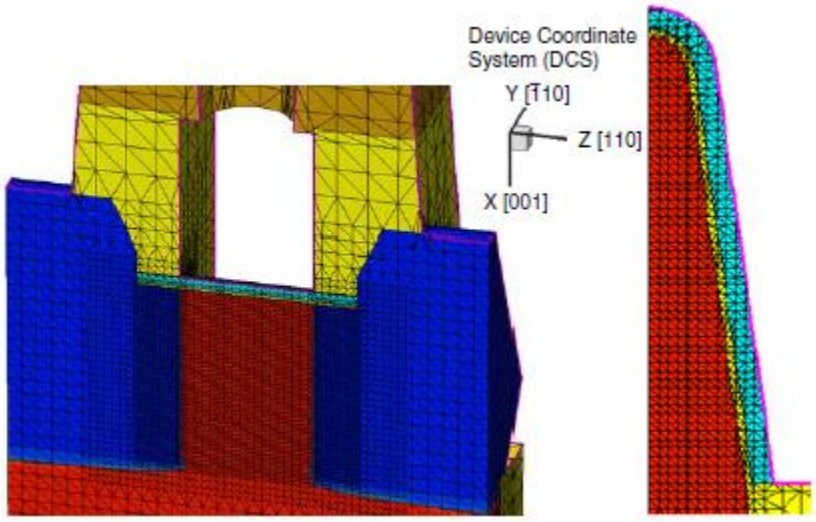


Figure 2-9: Meshing and Orientation of the simulated FinFET structure, [Sentaurus Template, **[*81*]**]

We briefly describe the basic device models incorporated in the TCAD tool (Synopsys, Sentaurus). These models, except the MC, are based on considering the mobility and the saturation velocity of the drift diffusion model as fitting parameters, yet they can yield quite accurate results for nano-scale device simulation [*23*].

A. Monte Carlo method (MC)

Monte Carlo (MC) simulation technique is considered the most accurate technique for simulating the carrier transport phenomena in semiconductor devices on the semi-classical level [*24*]. The main idea of the MC simulation is tracking large number of particles each one represents an electron through its journey along the device, trajectory, under the influence of electric field and subject to random scattering events. These trajectories are governed by classical newton's law and the carrier dispersion relation. The duration of the electron's free flight before getting interrupted by a scattering event, the type of the scattering event, and the final state after the scattering event are all chosen

based on probabilistic distributions. Simulating large number of these trajectories can yield very good average values for important physical quantities that can describe the average behavior of the carrier through the device and results a carrier distribution satisfies the Boltzmann Transport equation (BTE) [*25*]. Since the main MC algorithm is based on real physics, usually MC simulations are viewed as simulated experiments. In this investigation we take the MC results as the reference results when comparing the different models.

B. Modified drift-diffusion model (MDD)

Classical drift –diffusion (CDD) model works pretty well for long channel devices, where the transport is collision-dominated, however for short channel devices, the approximation of the transport as collision-dominated breaks down and near ballistic effects and strong velocity overshoot show up consequently the classical drift-diffusion model loses its validity. Due to the fact that the DD model is based on physics that can be derived from first moment of the Boltzmann transport equation (BTE) [*24*], it is not completely off what the actual model should be. So instead of going to sophisticated, time-consuming MC simulations, to some limits the classical DD model can be adjusted through modifying some parameters to fit the transport model in such small devices. To account for the velocity saturation effect, the mobility modeling is divided into two parts: a) low field mobility model, b) high field mobility model, which accounts for the velocity saturation effects. DD model incorporates a field dependent mobility model that provides smooth transition between low-field and high-field behavior. This model is called Caughey-Thomas (CT) field dependent mobility model and is expressed as:

$$\mu(E) = \frac{\mu_o}{[1 + (\frac{\mu_o E}{v_{SAT}})^{\beta}]^{1/\beta}} \qquad (2\text{-}5)$$

Where E is the lateral electric field (parallel to the oxide interface), vsat is the saturation velocity, β is the low field inversion layer mobility and β a constant.

The suitability of the DD model to simulate the short channel transistors through adjusting some parameters in the CT high filed velocity saturation model was first introduced in [*16*].

Then it was extended for the double-gate structures [*26*] through a fitting formula to define a length-dependent velocity saturation.

$$v_{sat}(L) = \frac{aL + b}{L + c} \qquad (2\text{-}6)$$

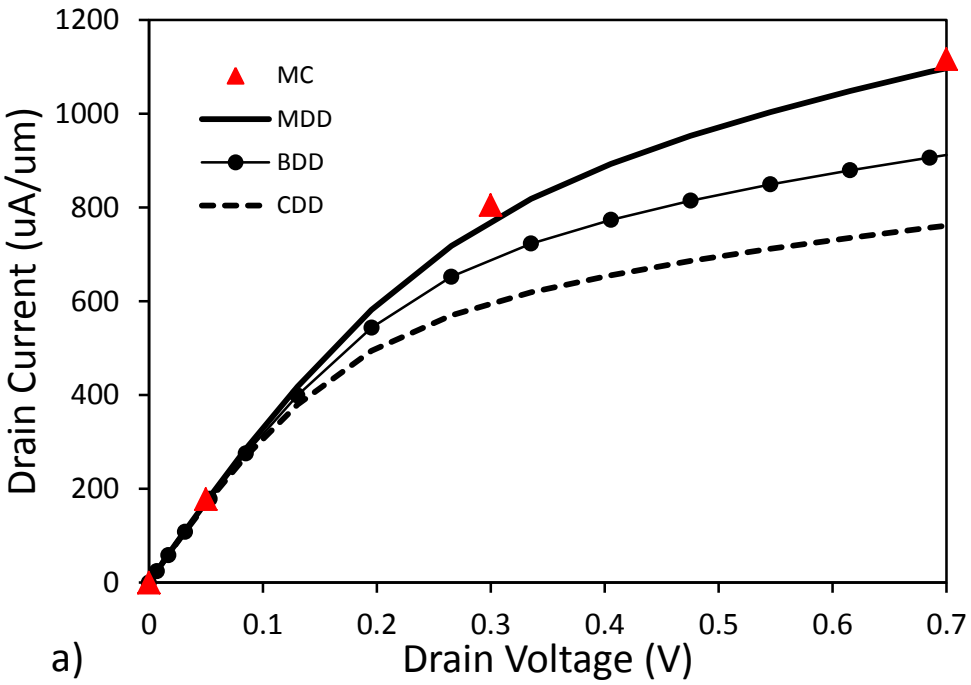
where a=1.5, b=21.6, and c=2.7 are fitting parameters to adjust the DD model for short channels simulations. For the triple-gate, we use this formula for the channel lengths of 17 nm and 15 nm to yield a velocity saturation values of 2.39×10^7 and 2.47×10^7cm/s respectively.

C. Drift-diffusion with Ballistic mobility model (BDD)

Adding more physically sounded adjustment to the traditional model can improve the results further. DD model is characterized by the mobility and the diffusion coefficient terms, consequently once one of these collision dominated transport related quantities loses its significance, the whole model fails. By re-examining the mobility term, according to [*27*], and [*28*], looking into the scattering current model expressed by the Landauer formula, a ballistic mobility like model was deduced which has a channel length dependence causing degradation of the mobility at short channel lengths. Then, a generic mobility model called apparent mobility was mathematically demonstrated taking the effect of the ballistic mobility into account and extends the mobility concept to very short channel lengths.

2.4.4. Simulation Results and discussion

As shown in Figure 2-10, output characteristics of two triple-gate structures; (a) for gate length of 17 nm, silicon fin thickness of 11 nm, (b) for gate length of 15 nm and silicon



fin thickness of 10 nm as projected by the ITRS for the years of 2015 and 2016, are simulated with all above mentioned models.

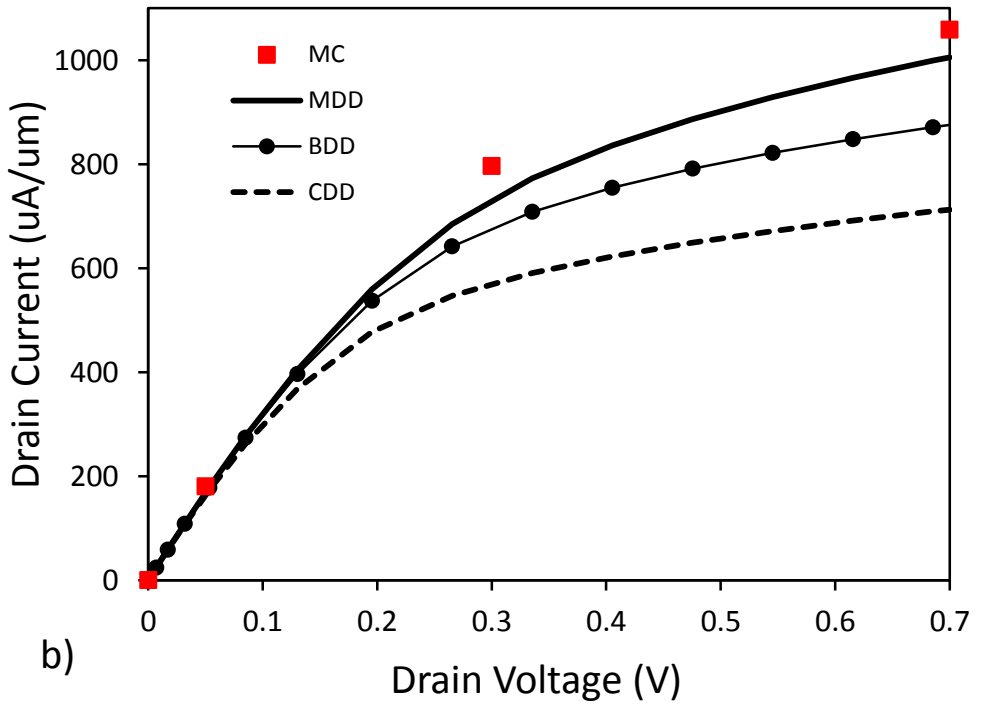


Figure 2-10: Output characteristics of Triple-gate FinFET simulated with Monte Carlo (MC), Modified drift-diffusion (MDD), Drift-diffusion with ballistic mobility model (BDD), and the classical drift-diffusion (CDD); (a) $\boldsymbol{L = 17\ nm, T_{si} = 11\ nm,\ H_{fin} = 27\ nm}$, (b) $\boldsymbol{L = 15.3\ nm, T_{si} = 10\ nm,\ H_{fin} = 27\ nm}$.

It is noteworthy to mention that realistic considerations are taken into account in these simulations such as a parasitic source/drain series resistance, strain effects, and high-k metal gate stack.

2.5. Conclusions

It is clear that the classical DD model underestimates the current characteristics as expected, and for the DD with ballistic mobility (BDD), it predicts the current in the linear region however the error increase as it goes deep into saturation, and it is clear that the modified DD model (MDD) is the most viable model and regenerated the MC results quite well, however it relies on non-physical fitting parameters formula to adjust for the length scaling effects which might vary from one structure to another. Accordingly we conclude that there is a need to conduct all the simulations based on MC approach to be able to draw appropriate conclusions.

3. A NUMERICAL STUDY OF NANO-SCALE TG-FINFET: 3D MONTE CARLO SIMULATIONS IN THE BALLISTIC AND Q-BALLISTIC REGIMES

In this chapter, nano-scale tri-gate (TG) FinFET with channel lengths down to 9.7 nm as projected by the 2013 International Technology Roadmap of Semiconductors (ITRS-2013) are simulated by means of quantum corrected 3-D Monte Carlo technique in the ballistic and quasi-ballistic regimes. Ballisticity ratio (BR) is extracted and found to reach values as high as 90% at LG = 9.7 nm. The impact of the ITRS-2013 scaling strategy on the BR, and ON-/OFF-states is discussed. Forward and backward electron velocity components are extracted along the channel to analyze the electron transport in detail. Velocity profile is found to be characterized by two critical points along the channel; each is associated with a change in the electron acceleration showing the physical significance of the off-equilibrium transport with scaling the channel length.

3.1. Introduction

Modeling and even simulation of nano-scale FinFET is a formidable challenge due to several factors. First, peculiar effects start to show up at these extremely scaled dimensions on the transport level such as hot electrons, velocity overshoot, ballistic and quasi-ballistic transport [*29*], [*10*], [*11*], [*30*] hence a careful treatment for the electronic transport must be considered. Second, with scaling the fin thickness, incorporation of quantum effects in the transport model is essential, hence quantum corrections are inevitable for proper accounting of the device electrostatics [*31*]. Third, the 3-D geometry of such non-planar multi-gate devices imposes new challenges especially on the computational level.

Therefore, choosing the correct transport model is considered the most serious challenge in the simulation of nano-scale transistors.

Several approaches have been proposed to account for such ballistic effects and strong off-equilibrium transport and even to determine the ultimate ballistic limit which is set by thermal injection from the source end [*11*], [*30*].

However, due to the lack of well-selected experiments which can appropriately discriminate between the various physical effects that interact with each other and suits such inextricable nature of electron transport on the nano-scale, they are still in need for more rigorous verification [*32*]. Therefore, more computationally intensive device models are required to study the transport in nano-scale devices such as Monte Carlo (MC) technique [*33*]. MC is considered the most efficient technique for simulating the carrier transport that involves hot electron phenomena and ballistic effects in semiconductor devices on the semi-classical level [*34*].

Previous works have been done based on 2-D MC simulations to study the ballistic and quasi-ballistic transport theories for nano-scale bulk and double gate (DG) SOI MOSFETs [*35*], [*36*], [*37*], [*38*], in addition to assessing the validity of the well-known analytical models developed in [*11*], [*30*]. In [*39*], MC simulations were used to verify newly developed backscattering models for Bulk MOSFETs within the Landauer theory. In [*40*], the same technique was used to study the scattering effects along the channel in DG MOSFETs, and further to get more insight about the main behavior of quasi-ballistic transport and determine the crucial parts of the channel that have the most contribution in limiting the ballistic transport. The scattering was turned on along some portions of the channel, and off in other parts and the different cases were compared. Some of the electron transport quantities have been discussed also by means of 2-D self-consistent MC simulations, where the evolution of the velocity distribution along the channel was analyzed [*36*].

Most of these computational studies confirm the general framework proposed in [*11*], [*30*], while others suggest additional complexities [*41*], [*42*], [*43*], e.g. the non-equivalence of the forward and backward velocities at the top of the barrier. However, due to the complexity of the 3-D nature of the tri-gate FinFET in addition to the

sophistication of the MC technique in 3-D, most of the work done for tri-gate FinFETs was for relatively long channels and based on conventional transport theories to develop analytical models [*44*], [*45*]. Little studies have been done for the tri-gate structure on the nano-scale based on 3-D Monte Carlo simulations [*46*], [*47*], and none of them provided a detailed study for the ballistic and quasi-ballistic transport in such devices, which is targeted in this work.

To keep short channel effects (SCEs) under control ($I_{OFF} = 100\ nA/\mu m$), ITRS implies scaling of the supply voltage V_{DD}, gate oxide thickness t_{ox}, and the fin thickness T_{fin}. For all the simulated devices, the same scaling strategy is mostly adopted as reported in Table.1.

Table 3-1: The main parameters of the simulated device

Year	2013	2015	2017	2019	2021
Physical Gate Length (nm)	20	16.7	13.9	11.6	9.7
Body Thickness (nm)	6.4	5.3	4.4	3.7	3.1
Fin Height (nm)	20.0				
Supply Voltage (V)	0.86	0.83	0.80	0.77	0.74
EOT (nm)	0.7	0.67	0.64	0.61	0.56
EPS_HIGH-K	22.0				
$t_{ox,ph}$ (nm)	2.56	2.53	2.46	2.42	2.37
Work Function (eV)	4.25				
Wafer/Channel Direction	001/100				
Channel Doping (cm^{-3})	$1x10^{15}$				
S/D Doping / S/D Ext. (cm^{-3})	$1.5x10^{20}$/ $1.5x10^{20}$				
S/D Extension (nm)	8.0				

3.2. Device Design and Simulation Methodology

Figure **3-1** shows the simulated structure under study, the channel length varies from 15.2 nm down to 9.7 nm, as projected by the 2013 ITRS. The work function is set to 4.25 eV. The fin top corners are rounded, close to the industry standard. Also the gate stack with nitride spacer formation is used. Source/drain (S/D) extensions are employed and set to 8 nm for all the simulated devices. In this work, we didn't consider the series resistance effect. The overlap distance between the gate and the S/D extensions (where S/D doping drops to 1×10^{19} cm^{-3}) is 1 nm. The mechanical stress is considered through all the simulation results. The scaling strategies as specified by the ITRS are mostly adopted in all the simulated devices. The main device parameters are reported in Table I and the corresponding doping profiles are shown in Figure **3-2**.

The simulation methodology is based on 3-D Monte Carlo simulation [*48*], incorporating quantum corrections using modified permittivity and work function taking into account the orientation dependence of the surface mobility [*47*], for both the ballistic and quasi-ballistic transport regimes. For the quasi-ballistic regime, scattering mechanisms include ionized impurity scattering, phonon scattering, and surface roughness scattering. For the ballistic regime, all the scattering mechanisms are switched off inside the channel volume such that the electrons have the full opportunity to transit from the source to the drain without encountering a single scattering event. Forward and backward components of average transport quantities such as the electron velocity are analyzed in the ballistic and quasi-ballistic transport. To do that, scalar product between the group velocity (with which the electron is directed) and a vector in the system coordinates directed towards the direction of transport [*48*] is evaluated (assuming the forward direction to be from the source to the drain). All Monte Carlo simulations are performed at 300K temperature.

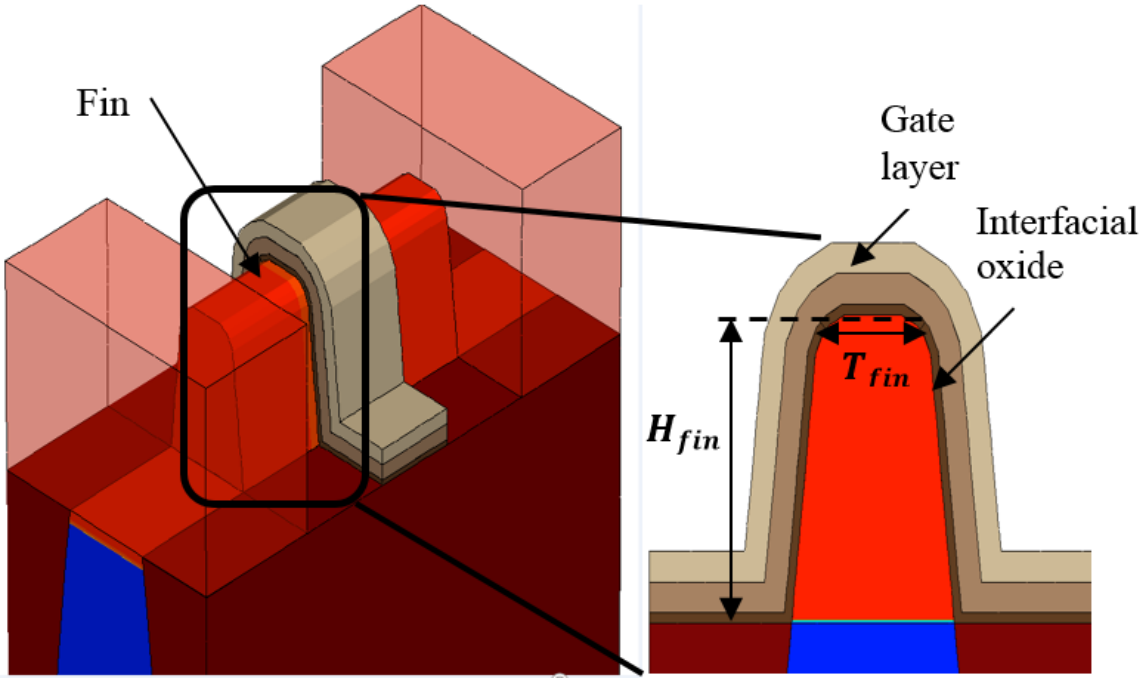


Figure 3-1: 3-D and 2-D representations of Tri-gate FinFET structure under study

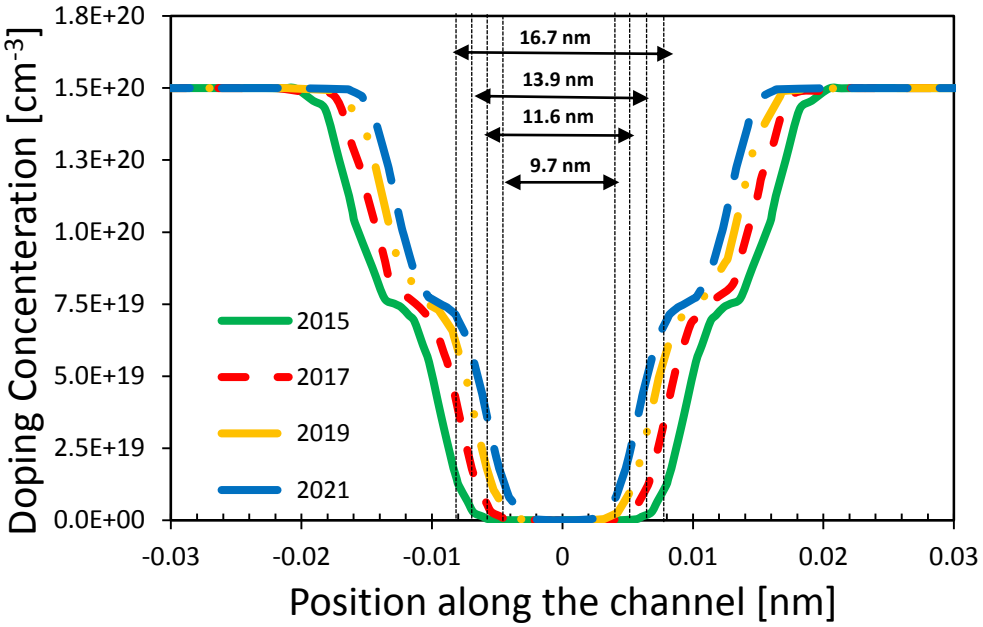


Figure 3-2: Doping profiles in cross sections of the simulated Tri-gate FinFET for channel lengths of 16.7, 13.9, 11.6, and 9.7 nm as projected to the years 2015, 2017, 2019, and 2021 respectively.

3.3. Simulations Results

3.3.1. Performance metrics with scaling

A. Off-state behavior:

The fundamental challenge in shrinking the transistor's gate length is to control the SCEs. Indeed, this problem is exacerbated in a nonlinear sense with approaching the 10-nm length. Figure 3-3 shows the behavior of the SCEs with length scaling normalized to the values at L_{ch} = 16.7 nm (130-mV/V drain induced barrier lowering (DIBL) and 130-nA/μm IOFF). The OFF-state behavior is studied by MC method, however, the Drift diffusion (DD) can be also used as an approximation. It can be noticed that scaling beyond the 13.9nm channel length yields a significant increase in I_{OFF} (exponential increase with scaling) and DIBL. Although the ITRS scaling strategy manages to keep the N_{inv} and V_{inj} almost independent of L_{ch} keeping the necessary assumptions of the ballistic theory.

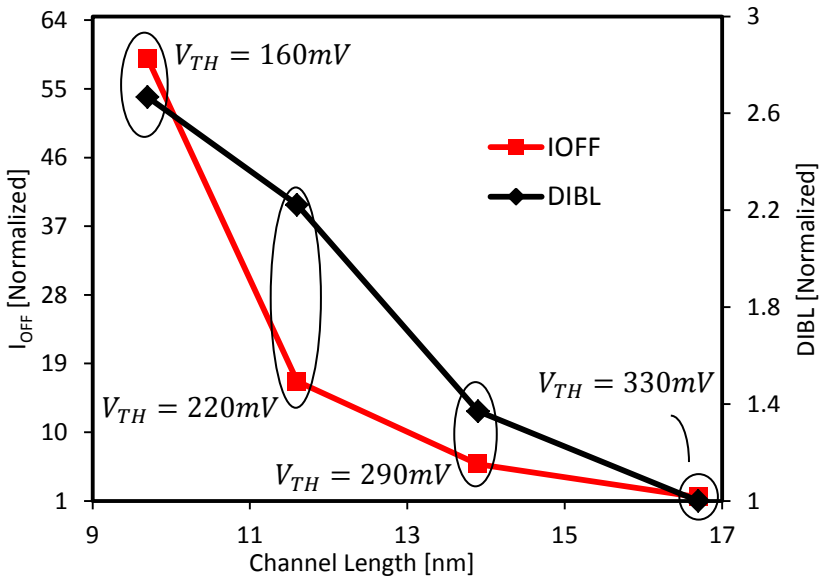


Figure 3-3: SCEs behavior of Tri-gate FinFET at different channel lengths showing the threshold voltage roll-off and the degradation of both DIBL and leakage current (IOFF) based on the adopted scaling strategy normalized to values at 16.7nm

According to our simulations, Fig. 3-3, it fails in keeping control of the SCEs, as shown in Figure 3-3. This from one side elucidates the severe degradation of the performance of such devices with down scaling and raises the need for other quick viable alternatives to extend the technology scaling. On the other side, such non-ideal effects raise additional complexities to the performance evaluation and require careful treatment in parameters extraction. For example, having a substantial DIBL leads to differentiation between the virtual source point and the top of the barrier (ToB) point, as discussed in [*49*], which used to refer to the same point interchangeably for an electrostatically well-tempered device.

B. ON-State Behavior

ON-current (I_{ON}) is considered the most indicative factor in evaluating the transistor's performance. This can be approached in two ways. First, I_{ON} per unit width, which is the most widely used metric to compare different devices. Second, we can still be concerned about the I_{ON} per device.

In this section, we consider both metrics separately. Approaching the ballistic transport has been considered to be the peak performance a device could ever achieve in terms of ION. Since operating in ballistic regime improves scaling benefits in a sense that:

1) Speed increases as a result of the transport in shorter channels, and
2) For these shorter channels being ballistic, i.e., comparable with the mean free path, is even better for the performance, since the electron transport would encounter less amount of scatterings yielding enhanced mobility, hence higher I_{ON}.

However, as a result of the necessity to adopt a scaling strategy to compensate the increase of the SCEs that involves scaling of other geometrical parameters besides the channel length, leading to reduced effective channel width, these benefits start to be undermined.

As shown in Figure 3-4, although a consistent improvement in ballistic ratios might be achieved with scaling (as discussed in Section IV), the devices are not able to attain similar performance improvements. Figure 3-4 (a) shows ION per unit width with technology scaling. The relative improvement (with respect to the preceding node) is decreasing with L_{ch} and almost saturates at 11.6 nm. Figure 3-4 (b) shows the corresponding ION per device with technology scaling. The relative improvement is also decreasing with the channel length. It does not saturate, but it diminishes instead, which means that the current itself saturates but not the improvement.

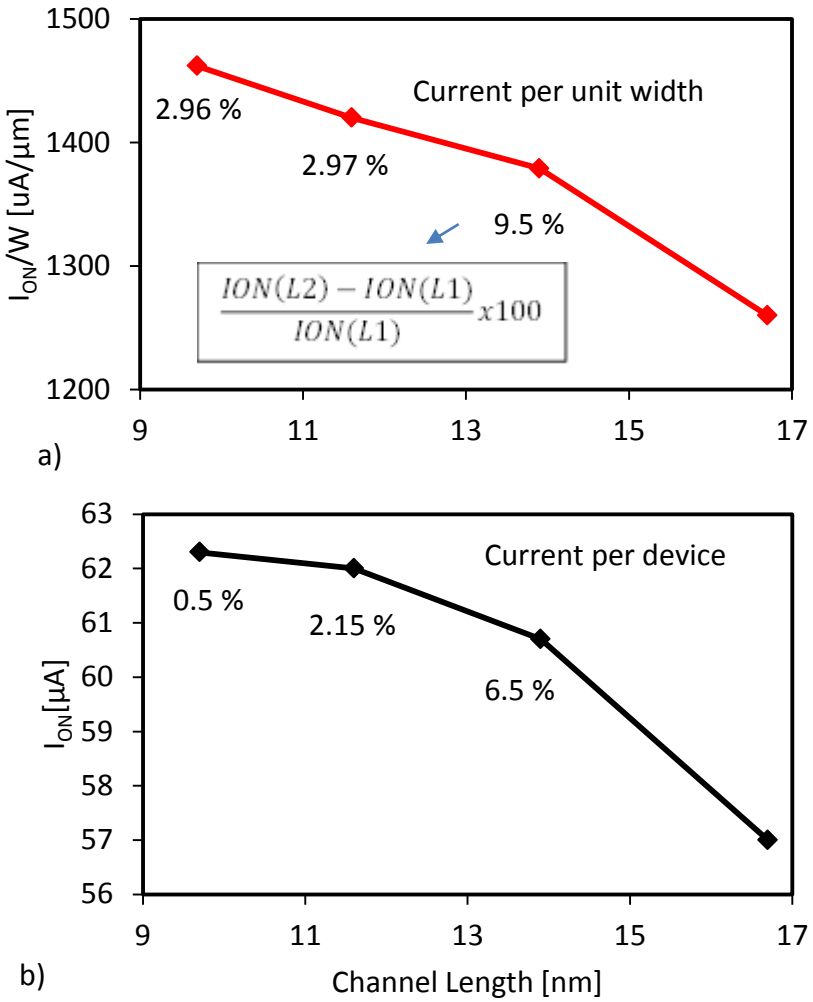


Figure 3-4: The behavior of the output current at VGS=VDS=VDD showing the relative improvement with technology scaling a) The current per device (corresponding to a device effective width; $\boldsymbol{W_{eff} = 2H_{fin} + T_{fin}}$), b) The device current per unit width.

Therefore, whether the relative improvement or the current itself saturates, this represents a serious slowdown in TG-FinFET scaling performance. In the latest ITRS edition, similar analogy has been pointed out. Along the roadmap, ION per unit width has been noticed for the first time to drop with technology scaling keeping fixed IOFF (controlled SCEs). This can have serious implications on advanced circuit design, as discussed in [*50*].

3.3.2. Ballisticity Ratio (BR): How close to the ballistic limit?

One of the most phenomenal questions is how close the technology scaling drives the transistors into the ballistic regime. This issue has been addressed before many times [*51*], [*52*], however it needs to be re-examined as the technology further scales and new devices emerge.

An aspect of special concern in exploring the carrier transport of nano-FETs and evaluating their performance; is the ballisticity ratio BR (related to the backscattering coefficient).

Several methods have been proposed for BR extraction [*53*], [*54*]; however whether they are based on experimental approachs or theoretical models, they usually encounter a number of assumptions and theoritical approximations that have been argued to be controversial [*53*]. For example, one of the most widely used techniques is based on the temperature dependence of I_{SAT} to extract the ratio of the mean free path to the cirtical length of the KT layer, hence BR [*54*]. However it turned out to be quite contraversial as discussed in [*53*], [*55*]. On the other side, being experimental does not guarantee the absolute validity of the extracted values, since the whole problem lies in the extraction of I_{BAL} value which remains a "theoretical" term. As a result, looking at the reported BR values in the literature a wide distraction is found. For example, previous works claimed that Si MOSFETs would opearate at a 50% BR [*52*], [*42*] regardless of L_{ch}. Others predicted a 65% BR at 10nm based on extrapollation of experimental results [*54*]. For NWFETs, based on rigorous analytic solution of the BTE, 75% BR was reported [*56*].

Even values as high as 90% were reported for junction less NWs of 20nm L_{ch} [*56*]. One benefit of our study, is to provide a reference up to date to compare against in adjusting newly developed experimental techniques for ballisticity extraction.

In this study we extract BR (defined as I_{Ball}/I_{QBall} at the on-state) at different channel lengths. To calculate I_{Ball} , all the scattering mechanisms are switched off inside the channel, while I_{QBall} considers all the scattering mechanisms. As shown in Figure **3-5**.a), I_{Ball} keeps almost a constant value, as expected to be a function of the device architecture only and independent of the channel length. Yet, the scaling strategy implies not only scaling of L_{ch} , but also the supply voltage V_{DD}, gate oxide thickness t_{ox}, and the fin thickness T_{fin}. According to the ballistic theory [*30*], [*11*], a constant upper limit for the on-current is resulted under the assumptions: a) the charges at the top of the barrier (ToB) N_{inv} is solely controlled by the gate, and b) the injection velocity V_{inj}, also at the ToB, is constant and independent of L_{ch}. For an electrostatically well-tempered device, those assumptions are achieved. However having substantial SCEs such as DIBL, would essentially affect the charges at the ToB. For example, at high V_{DS}, the ToB moves closer into the source/channel junction where a significant amount of charges pre-exist not induced by the gate, which in turn, affects both N_{inv} and V_{inj}values as discussed in [*49*]. Therefore, scaling of V_{DD}, t_{ox}, and T_{fin} efficiently suppresses such non-ideal effects on N_{inv} and V_{inj} and retrieves the ballistic limit. For the simulated devices, N_{inv} is found to be almost constant at around $(4.5 \times 10^{19} cm^{-3})$ and V_{inj} at around $(1.1 \times 10^{7}\ cm/s)$. Figure **3-5**.b) reports BR for the simulated devices, corresponding to the calculated currents in Figure **3-5**.a), along with the backscattering coefficient, defined as $r = (1 - BR)/(1 + BR)$. Longer channels are added to the study to notice the behavior of the ballisticity with length scaling, starting from 25 nm down to 9.7 nm. The following can be noticed; first, L_{ch} reduction of TG-FinFET yields consistent improvement of BR. Second, the increasing slope of the BR is quite small for the relatively longer channels, and turns much steeper at around 13.9 nm (this point will be further discussed in the subsequent sections). Moreover, BR is around 73 % for L_{ch}=25

nm and reaches values as high as 90 % at channel length of 9.7 nm. These high values, however, come at the expense of increasing the SCEs.

Although the ITRS scaling strategy manages to keep the N_{inv} and V_{inj} almost independent of the channel length keeping the necessary assumptions of the ballistic theory, according to our simulations, it fails in keeping control of the SCHs. In particular, at $L_G = 16.7, 13.9, 11.6,$ and $9.7\ nm$, the DIBL varies as 136, 186, 300, and 350mV/V, respectively. This from one side elucidate the severe degradation of the performance of such devices with scaling and raise the need for quick other viable alternatives to extend the technology scaling. In addition, this reassures the trade-off between the ballisticity and the SCEs. From the other side, these non-ideal effects raise additional complexities to the performance evaluation and require careful treatment in parameters extraction.

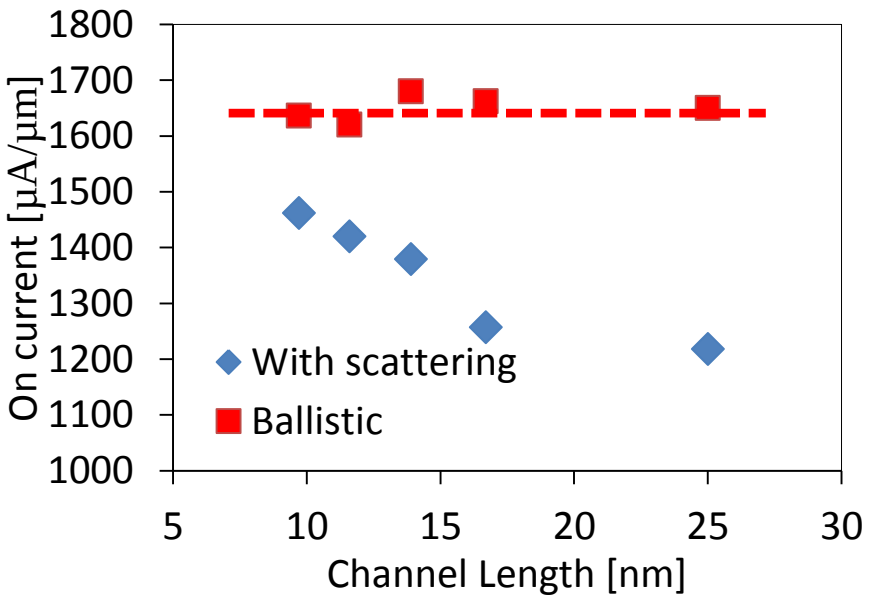


Figure 3-5: Drain current at VD=VG=Supply Voltage, normalized to the effective channel width, $\boldsymbol{W_{eff} = 2H_{fin} + T_{fin}}$, at scaled channel lengths, body thicknesses, supply voltages, and oxide thicknesses projected by the 2013 ITRS as reported in Table 1

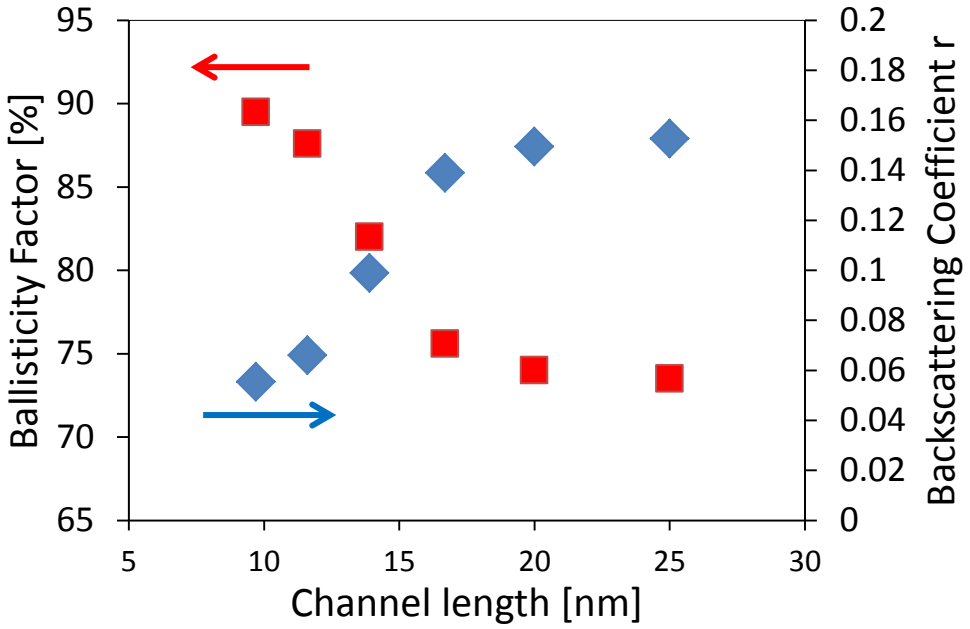


Figure 3-6: Ballisticity factor in the left y-axis and corresponding backscattering coefficient in the right y-axis, with scaling the channel length as reported in table 1, at body thickness = 4 nm, supply voltage = 0.78 V.

For example, having a substantial DIBL leads to differentiation between the virtual source VS point and the ToB point as argued in [*49*], which used to refer to the same point interchangeably for an electrostatically well-tempered device.

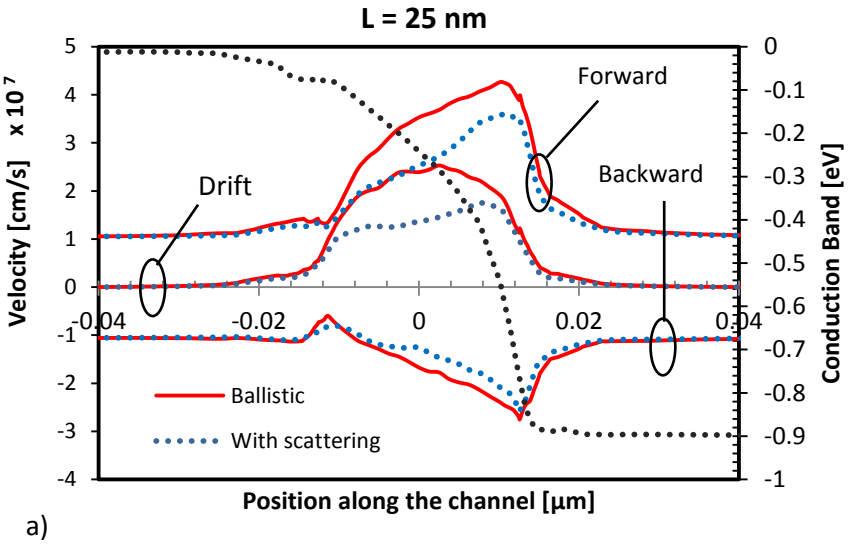
Regarding the ballisticity ratios, these achieved values, in fact, are quite different from other studies that were pessimistic about reaching such high values, claiming that the ballisticity appeared to saturate around ~ 50-60 % for Si devices as the channel length is aggressively reduced further [*42*], [*52*]. Moreover, it has been argued that such ballistic limit may not be achievable and suggested that the surface roughness scattering at the oxide interface is the main responsible for such limitation off the ballistic limit [*42*], [*52*]. However, this study shows that approaching the ballistic limit is not the main problem; other factors should be taken into consideration in judging the device performance.

3.3.3. Electron Velocity Evolution along the Channel

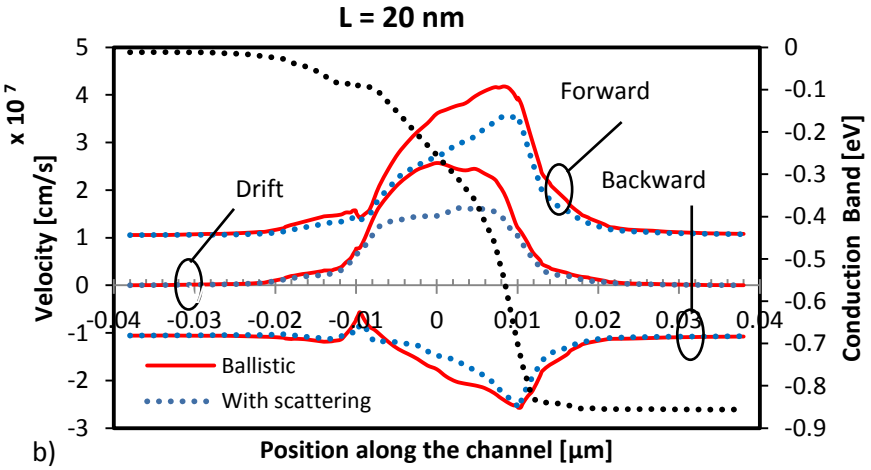
The ballistic transport theory was initially developed based on Landauer approach [*57*] which decomposes the electron transport quantities into forward and backward directed components with positive and negative group velocities, respectively. In this section, therefore, we adopt a direction-dependent analysis in which we analyze the electron velocities of the forward (v^+) and backward (v^-) fluxes separately, in addition to the average drift velocity (v_d).

Figure **3-7** a)-f) shows the evolution of v^+, v^- and v_d along the channel, and the conduction band profile for the ballistic and quasi-ballistic transport regimes with scaling the gate length from 25 nm down to 9.7 nm. All the simulated results are done for the on-state ($V_G = V_D = V_{DD}$). Consider first, the conduction band (CB) profile along the channel. We focus on two relevant phenomena, the first is a transport-related and the second is an electrostatics-related.

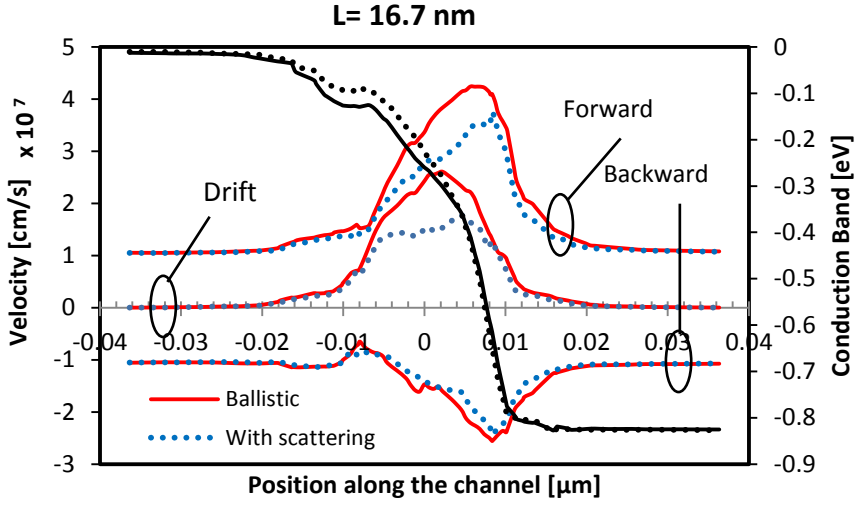
First, for the transport-related, at large drain bias for long channels, the electron charge density near the drain eventually falls to very low values leading to the so-called pinched-off region with very low conductance [*58*], [*59*]. Consequently, the electric field rises up to very high levels increasing the electron velocities along this pinched-off region keeping the current continuity. An inflection point appears in the CB profile referring to the beginning of such high field region. This used to be a characteristic for long channel transistors and was associated with velocity saturation. In our results, however, for the relatively longer channel lengths (~ 25 nm) as shown also in Figure **3-7**.a-d), the CB profiles show very like behavior with a clear inflection point, but instead it is associated with velocity overshoot not saturation. Note that, this phenomenon is common for both the ballistic and quasi-ballistic channels, thus the scattering processes have no effect in this case on the CB profile. However, it affects the velocity profiles differently in the ballistic and quasi-ballistic transport regimes as it will be discussed.


L = 25 nm
Velocity [cm/s] x 10 7
Conduction Band [eV]
Position along the channel [µm]
Forward
Backward
Drift
Ballistic
With scattering

a)


L = 20 nm
x 10 7
Velocity [cm/s]
Conduction Band [eV]
Position along the channel [µm]
Forward
Backward
Drift
Ballistic
With scattering

b)


L= 16.7 nm
Velocity [cm/s] x 10 7
Conduction Band [eV]
Position along the channel [µm]
Forward
Backward
Drift
Ballistic
With scattering

c)

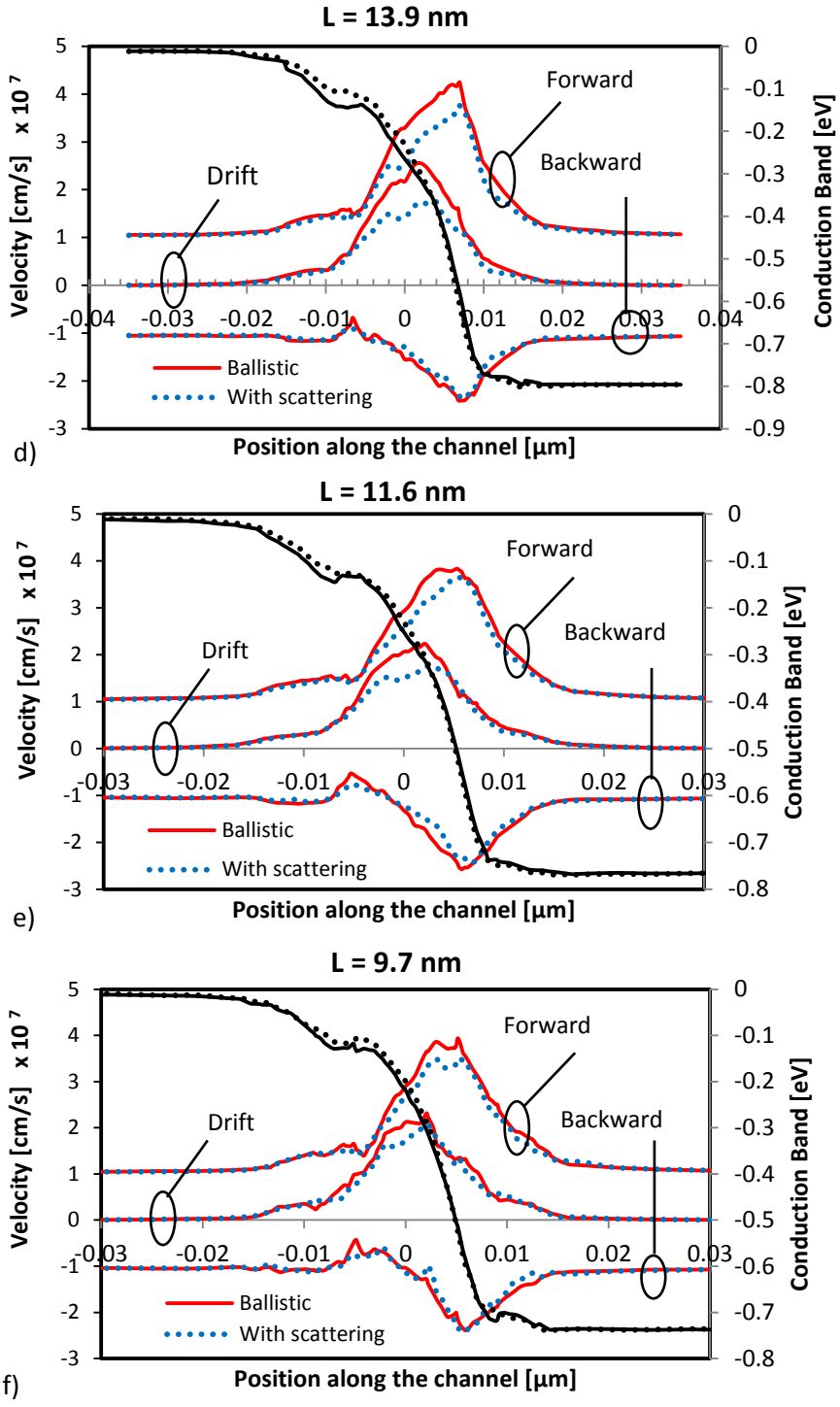


Figure 3-7: Different average electron velocity components: drift, forward, and backward; and the conduction band profile along the channel 15 $\boldsymbol{A^o}$ below the Si-SiO_2 interface at various channel lenghts: a) L=25 nm, b) L=20 nm, c) L=16.7 nm, d) L=13.9 nm, e) L=11.6 nm, f) L=9.7 nm. Solid lines: indicating the ballistic case, dotted lines: including all scattering mechanisims. All simulations are done for the on-state ($\boldsymbol{V_G = V_D = V_{DD}}$).

Second, for the electrostatics-related, in case of ballistic channel the top of the barrier is slightly lower than the quasi-ballistic case due to the increased backward flow of electrons. Consequently, the barrier height slightly floats up to compensate this effect as dictated by the electrostatics to keep almost constant inversion charge density at the top of the barrier [*30*]. However, the so-called virtual source point is almost unchanged in the two cases. The implications of the first phenomenon will be further discussed in the rest of this section.

Next, we discuss the evolution of the velocity components along the channel. For the forward and backward components, as it can be noticed also in , first the backward velocity components in the ballistic and quasi-ballistic regimes are very close to each other over all channel lengths, and as the channel length shrinks, they eventually coincides such that the deviation from the ballistic case is almost negligible. However, most of the velocity deviation lies in the forward components, and as the channel length shrinks this deviation gradually shrinks too.

Second, the velocity increases as the electrons move in the channel towards the drain but the v^+component is clearly exceeding the v^- in absolute values which is consistent with previous results observed by MC in [*37*]. As a result the forward component has the dominant effect on the overall average drift velocity. Third, clear overshoot is observed in the velocity profiles, especially in the v^+ component. Note that, here we define the overshoot behavior as a sudden increase in the electron acceleration forwarding to the drain. In addition, for the quasi-ballistic case, in all channel lengths, the velocity profiles almost keep the same overshoot peak value which is about $(3.6 \times 10^7 cm/sec)$. While in the ballistic case, they exhibit a different behavior, a slight decrease in the overshoot peak velocities is observed with shrinking channel length starting from around $(4.1 \times 10^7 cm/sec)$ at 25 nm to around $(3.75 \times 10^7\ cm/sec)$ at 9.7 nm. This decrease in the peak velocity value in the ballistic case had been observed before with MC simulations in [*41*] without emphasis.

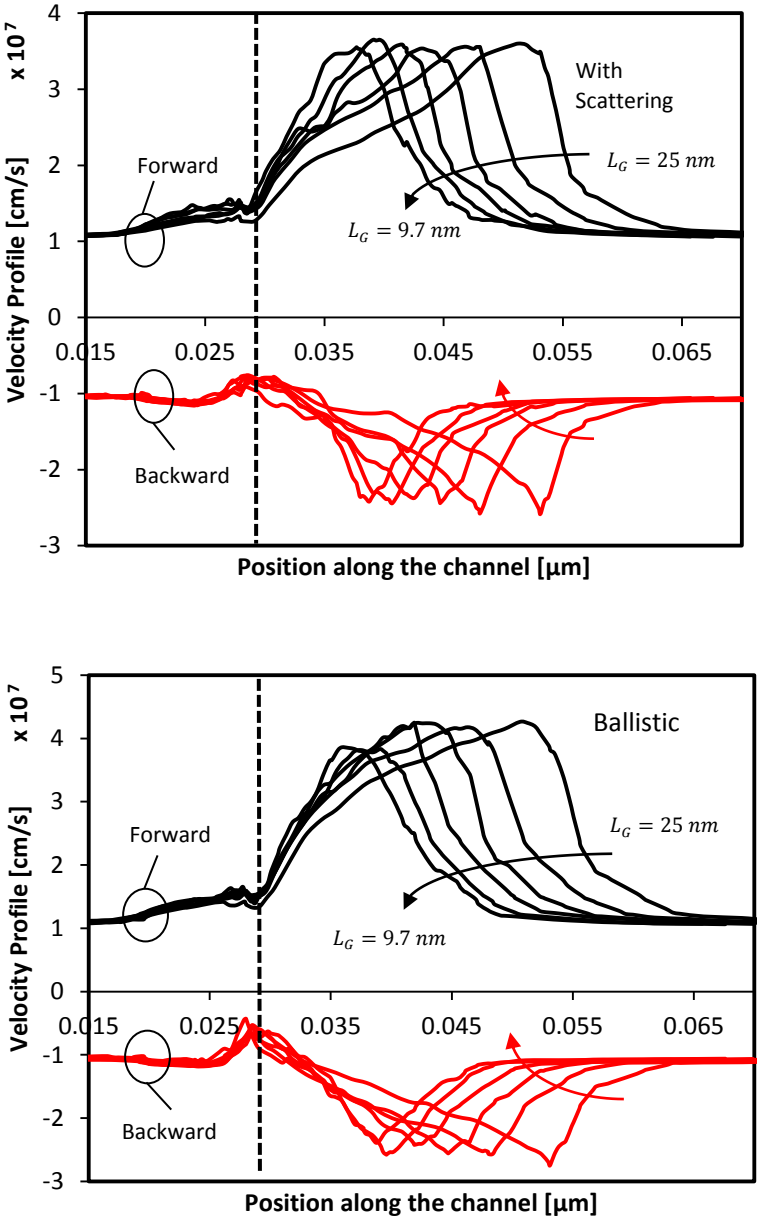


Figure 3-8: Forward and backward components of the electron velocity for Tri-gate FinFET 15 A^{o} below the Si-SiO2 interface at various channel lengths 25, 20, 16.7, 13.9, 11.6, 9.7 nm; a) For the Quasi-ballistic case (including scattering), b) For the ballistic case.

Next, we study the influence of the channel position in limiting the ballistic transport. This point has been studied before in [*40*] by performing three types of simulations, first applying scattering in the first half of the channel only, then in the second half only, and finally comparing with full ballistic channel so as to determine the impact of each part separately. In [*37*] the scattering events contributing to the backward current flux are computed along the channel, and their decay length was extracted and compared to the KT-length as defined in [*30*]. In this work, though, we investigate the same point using 3-D Monte Carlo simulations but from a different point of view. We analyze the velocity profiles with position along the channel in both the ballistic and quasi-ballistic regimes to get insight about the most crucial part of the channel.

We define a characteristic for each half of the channel. For the first half, it is characterized by the initial electron acceleration by which the electrons step up with at the beginning of the channel from the source side, and for the second half, it is characterized by the overshoot behavior as a result of the off-equilibrium transport in the high field region.

As shown Figure 3-8, for the first half, that is the beginning of the channel around the top of the barrier, we find that the forward electron velocity v^+ has less initial acceleration in the quasi-ballistic than in the ballistic case, which leads to less percentage velocity of the ballistic limit. Then, looking at the second half, around the high field region near the drain end, it can be noticed that the velocity overshoot is much apparent for the quasi-ballistic than the ballistic case. Although it ends up with lower peak velocities, the quasi-ballistic forward flow of electrons are much accelerated, as a result of the overshoot, leading to higher percentage velocity of the ballistic along this spatial segment.

From the above observations we can conclude the following: the first half of the channel, which is associated with a near-equilibrium transport, is less ballistic leading to significant deviation from the ballistic regime. While the second half is much closer to the ballistic transport regime. This can be attributed to the strong off-equilibrium transport in the second half of the channel near the drain end.

As the channel length shrinks, the overshoot behavior dominates the velocity profile along the channel. In other words, the transport becomes off-equilibrium over a wider portion of the channel leading to the increase of the ballisticity factor. This is also consistent with the appearance of the aforementioned inflection point in the CB profile for the relatively long channels and gradually becomes indistinguishable as the channel length shrinks. Hence, we can say that the off-equilibrium phenomena make the device more ballistic. And as a consequence of the continuous shrinking of channel length and increasing variations of the electric filed on the spatial and time scales, the electron transport eventually becomes more off-equilibrium, consequently the devices are expected to be more and more ballistic.

3.3.4. Discussion

Further examination for the results reveals that there are basically two points along the channel: X, Y, depicted in Figure 3-9, where the electron transport changes substantially. First we start analysis for the long channel case. At the beginning of the channel (where the ToB roughly lies) electrons are thermally injected with a certain initial acceleration. For the ballistic regime, this acceleration is almost constant and independent of the channel length (Figure 3-8). However, for the q-ballistic, it increases as the gate length shrinks approaching its ballistic limit. As the electrons go in the channel, reaching point X, the electron acceleration obviously changes. This behavior was previously observed in [*35*], [*33*]. An early explanation was presented in [*11*] suggesting that this point marks a transition from thermal injection, as coming from the source, and off-equilibrium injection in the channel. Accordingly, X is expected to be located right at the ToB, however our MC results show that point X is located few K_BT eV energy downside. The second point is Y where the velocity profile exhibits another change in the electron acceleration. For the q-ballistic regime, a steep increase in the electron acceleration occurs indicating strong velocity overshoot.

This is expected because the velocity overshoot by definition is due to the nonequivalence of electron energy and momentum relaxation times as resulted from the different scattering processes [*4*], and the ballistic transport by definition is free of any scattering processes. This point also indicates the beginning of the off-equilibrium region which is generally characterized by the overshoot behavior. Therefore one might expect that this overshoot should start right at the inflection point on the CB profile which indicates the beginning of such off-equilibrium region as discussed above. However, as shown in Figure 3-9.b) on the left, the overshoot starts a bit earlier showing anticipating like behavior.

For the short channel, Figure 3-9. a, b) on the right, the regions indicated by points X and Y overlap and the two points reduce to a single point due to the increased off-equilibrium regime with scaling the channel length. This analysis shows the evolution of the electron velocity along the channel, indicating the profound impact of the off-equilibrium regime on the electron transport as the gate length shrinks.

Finally, looking again at the BR curve with scaling (Figure **3-6**.b)); initially BR increases slowly with shrinking L_{ch}, during this scaling range, the high field region is quite far from the ToB. With further scaling, the high field region gets closer to the ToB.

Once the difference between the ToB and the inflection point on the CB profile is around 8.5 KT or less, the BR is significantly impacted exhibiting a sudden increase before it slows down again approaching 90% when the two regions eventually merge and points X and Y coincide. This can explain the BR behavior with scaling.

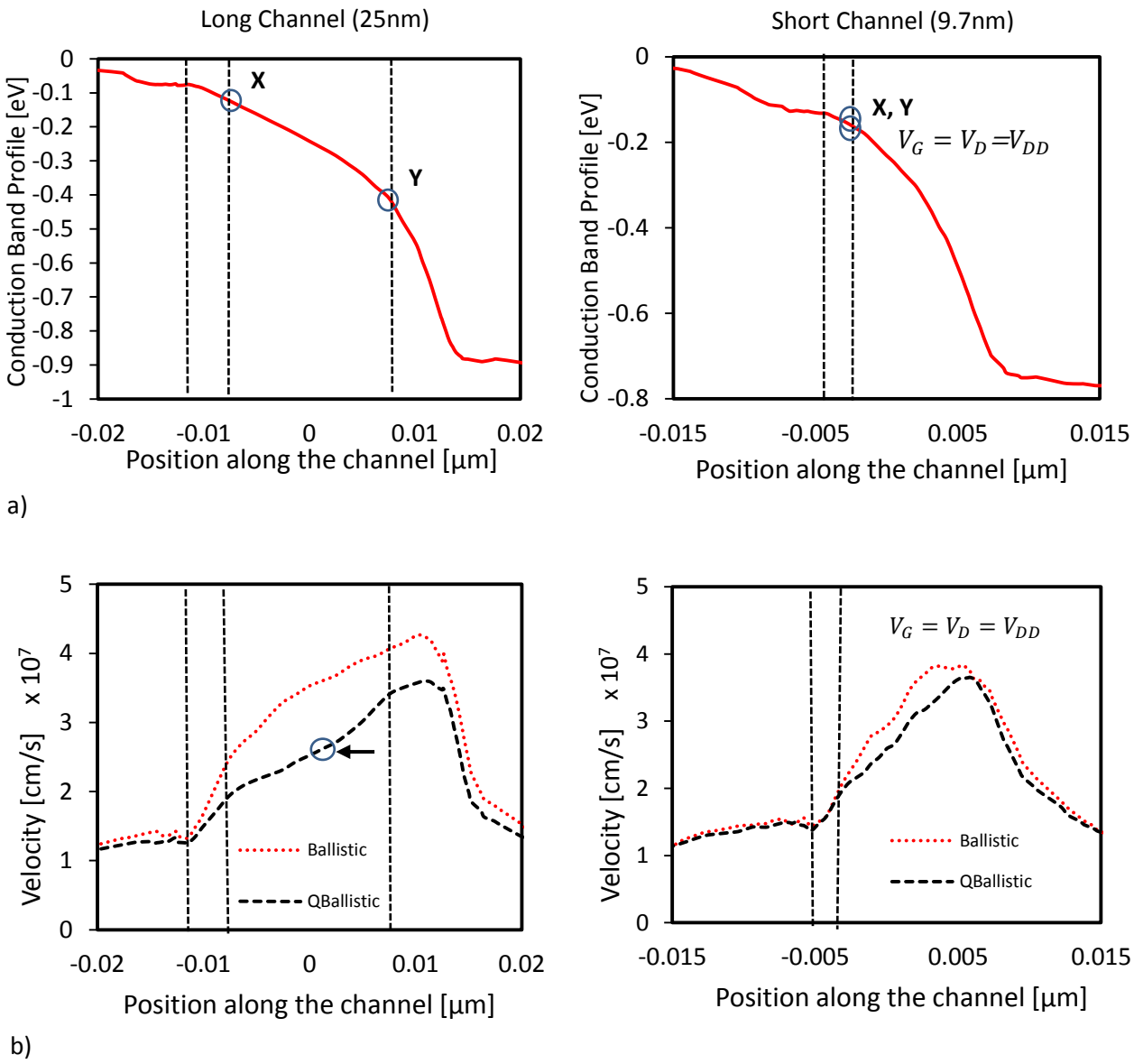


Figure 3-9: a) CB profile along the channel for long and short channels, b) Velocity profiles for long and short channels.

3.4. Conclusions

Quantum-corrected 3-D Monte Carlo simulations demonstrate that channel length reduction of TG-FinFET yields consistent improvement of the ballisticity factor reaching values as high as 90 % at channel length of 9.7 nm.

Despite the reported improvement in the ballisticity with scaling, the devices are not able to attain expected performance improvements. This can be seen in (i) the diminishing improvement of the on-current with scaling the channel length of TG-FinFET, (ii) the increasing SCEs specially starting from 11.6 nm channel length despite the adopted scaling strategy.

The simulation results reveal that the electron transport along the channel is characterized by two critical points dividing the velocity/CB profiles into three spatial regions with significant change in the electron acceleration/energy. However, as the gate length shrinks the two points reduce to a single point; hence the velocity profile is characterized by two regions instead of three as a result of the dominance of the off-equilibrium phenomena with extreme length scaling.

The BR behavior with scaling was analyzed. It was found that initially BR increases slowly with shrinking L, during this scaling range, the high field region is quite far from the ToB. With further scaling, the high field region gets closer and closer to the ToB. Once the difference between the ToB and the inflection point on the CB profile is around 8.5 KT or less, the BR is significantly impacted exhibiting a sudden increase before it slows down again approaching 90% when eventually the two regions merge and points X, Y coincide.

4. EVALUATION OF TG-FINFET SCALING ROADMAP IN CIRCUIT DESIGN

4.1. Introduction

Being in the time where all of the major foundries have announced FinFET technologies for their most advanced processes. Intel introduced the 1st generation TG FinFET for the 22 nm node, and 14 nm as the 2nd generation, Figure 4-1, also TSMC for their 16 nm process, and Global foundries and Samsung for their 14 nm processes. At the same time, extensive research and development are carried out everywhere for the upcoming nodes down to 7 nm, Figure **4-2**. As with any new process technology, the ultimate goal is to enable efficient circuit design that is incorporated in diverse applications. And a successful process should yield three main aspects: a) higher performance, b) lower power, and c) lower cost per transistor, and that is what Moore's law is all about, Figure **4-3**.

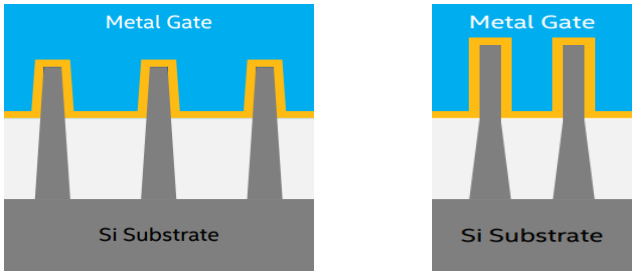


Figure 4-1: TG-FinFET a) 22 nm 1st Generation, b) 14 nm 2nd Generation Tri-gate Transistor [INTEL presentations]

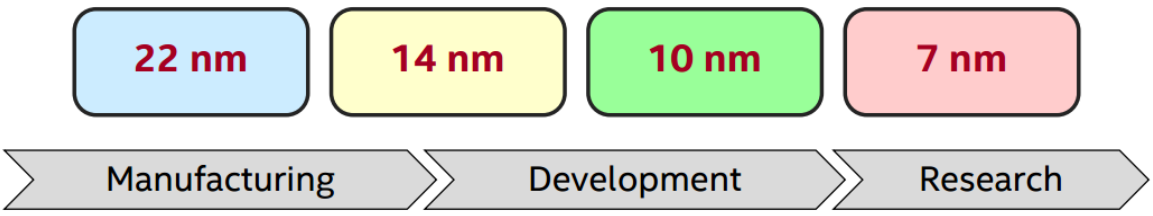


Figure 4-2: FinFET demonstration road map

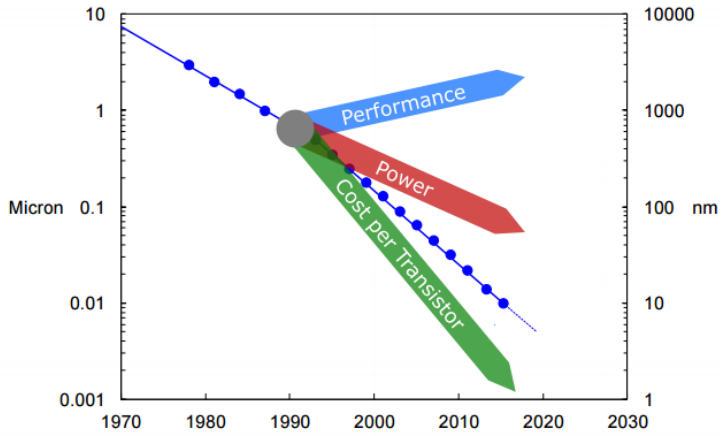


Figure 4-3: Transistor scaling guidelines for circuit design [Intel presentations]

However to which extent TG-FinFET will be scalable is still questionable. For Intel producing its second generation of TG-FinFET applying the scaling procedure for the first time of such devices, they already introduced a bunch of modifications to achieve the targeted performance. By looking at Figure 4-1, the main process modifications encountered for the transition from 22 nm to 14nm can be illustrated. First, the Fin pitch has been reduced from 60 nm to 42 nm leading to tighter fin pitch to improve the density. In addition, the 14nm devices incorporate fewer number of fins which is basically to reduce the parasitic capacitance associated with these 3-D structures hence reduce the overall active power of the chip. Second, the Fin itself has been thinned and became taller. Being thin, improves the off-state behavior by improving the device electrostatics, and being tall (increasing the fin height from 34 nm to 42 nm), improves the on-state behavior since it effectively increases the device width which consequently leads to higher on-current.

In this study, we focus on TG-FinFET from circuit design and performance perspective. The objectives behind this work are to examine the ultimate scaling limits of FinFET devices and their potential strengths in circuit design by examining different process

nodes at different channel lengths. For this purpose, two different types of studies have been carried out:

a) Self-study: where the performance of TG-FinFETs is investigated over different technology nodes, scaling the channel length from 20, 16, 14, 10, and 7 nm. In this part, predicative technology models (PTM) [*60*] are used in the simulations.
b) Peer-to-peer study: where the performance of TG-FinFET is assessed with respect to the most recent commercial technologies. In this part, 28nm FD-SOI PDK has been used to compare against.

For both of the studies, SRAM memory cell has been used as a test vehicle to assess the performance as a basic building block and being of extreme importance in modern SoC applications. The first study is considered elementary study in terms of the simulations setups, concerned only with the very basic performance metrics of SRAM cells. However, the study is considered more advanced where more sophisticated simulations and characterizations has been considered.

4.2. Performance Evaluation of FinFET based SRAM under Statistical VT Variability

In this study, the performance of extremely scaled FinFET-based 256-bit (6T) SRAM is evaluated with technology scaling for channel lengths of 20nm down to 7nm showing the scaling trends of basic performance metrics. In addition, the impact of threshold voltage variations on the delay, power, and stability is reported considering die-to-die variations. Significant performance degradation is found starting from the 10nm channel length and continues down to 7nm.

Static Random Access Memory (SRAM) occupies a significant portion of all system-on-chips and microprocessors as an efficient embedded memory block [*61*]. As a result of the increasing demand for higher performance and integration, higher density SRAM cells are designed with the minimum size transistors in a given technology node.

Increased process variability and device reliability issues increase the necessity for performance evaluation of SRAM design methodologies and topologies with technology scaling.

On one hand, shrinking the channel length significantly increases the short channel effects (SCEs) which in turn degrade the basic cell metrics such as the leakage power.

On the other hand, emerging novel devices such as FinFETs poses new challenges by adding new variability sources such as the Fin thickness variations as a result of increased line edge roughness. In addition to new design issues such as width quantization which limits the design optimization [*62*].

However, having new geometry parameters such as the fin thickness, the quantized number of Fins, and even surface orientation opens the way for new design optimization techniques [*63*]. On top of the challenges of scaling of SRAM on the design level as specified by the ITRS-2013, is to maintain adequate noise margins and control key instabilities and soft-error rates in the presence of random threshold voltage (V_T) fluctuations. In addition to the difficulty in keeping the leakage current within tolerable limits.

Some studies have discussed the FinFET SRAM performance at the nano-scale. For instance, a simulation study for 14 nm SOI FinFET technology has been reported showing the impact of the relevant sources of variability and reliability on the cell stability [*64*], [*65*]. In [*63*], the FinFET SRAM design space is discussed, under different Fin thicknesses and Fin heights, to optimize stability, delays and leakage current but at constant channel length.

In this study, we report the conventional (6T) SRAM cell operation's limits within a given range of threshold voltage variations along with different technology nodes starting from 20 nm down to 7 nm.

Table 4-1: The simulated device parameters

Device	TG-FinFET				
L (nm)	20	16	14	10	7
T_{fin}(nm)	15	12	10	8	6.5
H_{fin} (nm)	28	26	23	21	18
V_{DD}(V)	0.9	0.85	0.8	0.75	0.7
Fin ratio (N_{fin}) (PU:PD:PG)	(1 : 3 : 2)				

The variations are considered die-to-die variations. The operation's limits are determined through the evaluation of read/write static noise margins (RSNM/WSNM) as an indication for the cell's stability, read and write delays, active and leakage powers.

4.2.1. Simulation Methodology

In this study, predictive technology model (PTM-MG) files [*60*] for Multi-gate devices (TG-FinFET in our case) are used from 20 nm down to 7 nm technology nodes for low-standby power devices (LSTP) with the BSIM-CMG compact models. A scaling strategy is adopted according to the PTM models which involves, besides the scaling of the channel length (L), scaling of the supply voltage (V_{DD}), fin thickness (T_{fin}), and fin height (H_{fin}). The used device parameters are reported in Table **4-1**. Tri-gate FinFET structure is used such that the effective channel width is ($W_{eff} = 2H_{fin} + T_{fin}$).

Regarding the performance metrics, the read delay is calculated as the time period from the 50 % point of the word line (WL) low-to-high transition to a 10 % difference point developed between the BL and BLB.

The read/write static noise margins (RSNM/WSNM) are used as a measure for the read/write operations stability of the SRAM cell respectively, and are defined as the maximum absolute DC voltage around the half-supply pre-charged bit-lines (BL,

BLB) that causes the stored state of the cell not to flip during the read operation, or the maximum absolute DC voltage below V_{DD} for BL and above '0' for BLB that changes the state of the cell for a successful write operation.

The leakage power consumption is calculated for the SRAM cells in the idle mode; when the access transistors are cut-off and the bit-lines are left floating.

4.2.2. Simulation Results and Discussions

A. Read/Write Delays

Read/ Write delays are key parameters in evaluating the performance of SRAM cell. Figure 4-4, Figure 4-5 show the sensitivity of the SRAM read/write delays with technology scaling and V_T variations (from - 40 % to 40 % of the nominal value). As it can be noticed, with increasing the threshold voltage the delay increases as a result of decreasing the overdrive voltage hence reducing the transistors' currents. For write operation, the delay encounters a variations of around +/- 25 % over the +/- 40 % threshold variations, and for the read operation, the variation in the delay is around +/- 35 % which is quite higher, with respect to the delay value at the nominal V_T.

On the other side, observing the behavior with the technology scaling. For the write operation, the delay is continuously decreasing with scaling down the technology as a result of shrinking the channel length despite the scaling of the supply voltage which usually leads to increase in the delay. However for the read delay it is quite different, since degradation is observed starting from the 10 nm node down to the 7 nm as it can be seen in the inset of Figure 4-5. To understand this behavior, we plot the read current at each technology node as shown in Figure **4-6**.

As it was discussed in the first section, with technology scaling, other parameters are scaled besides the scaling of the channel length such as the supply voltage, the fin thickness, and the fin height which is basically to compensate for the increased SCEs associated with such extreme scaling.

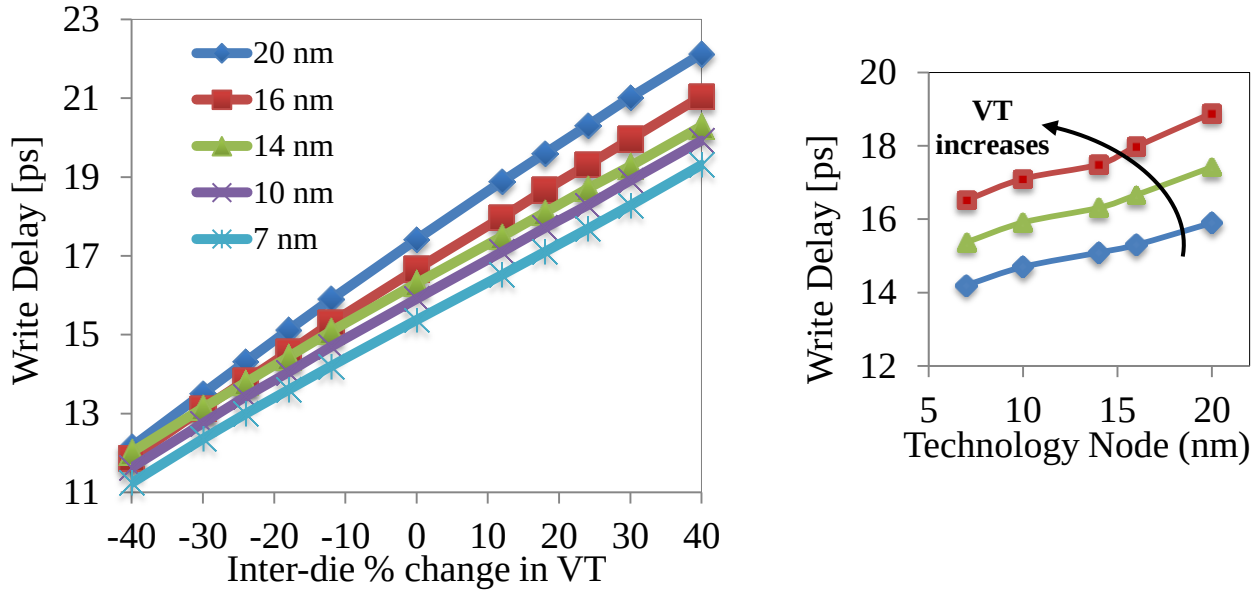


Figure 4-4: SRAM Write delay sensitivity to threshold voltage inter-die variations range of +/-40 % at various technology nodes from 20nm down to 7nm node.

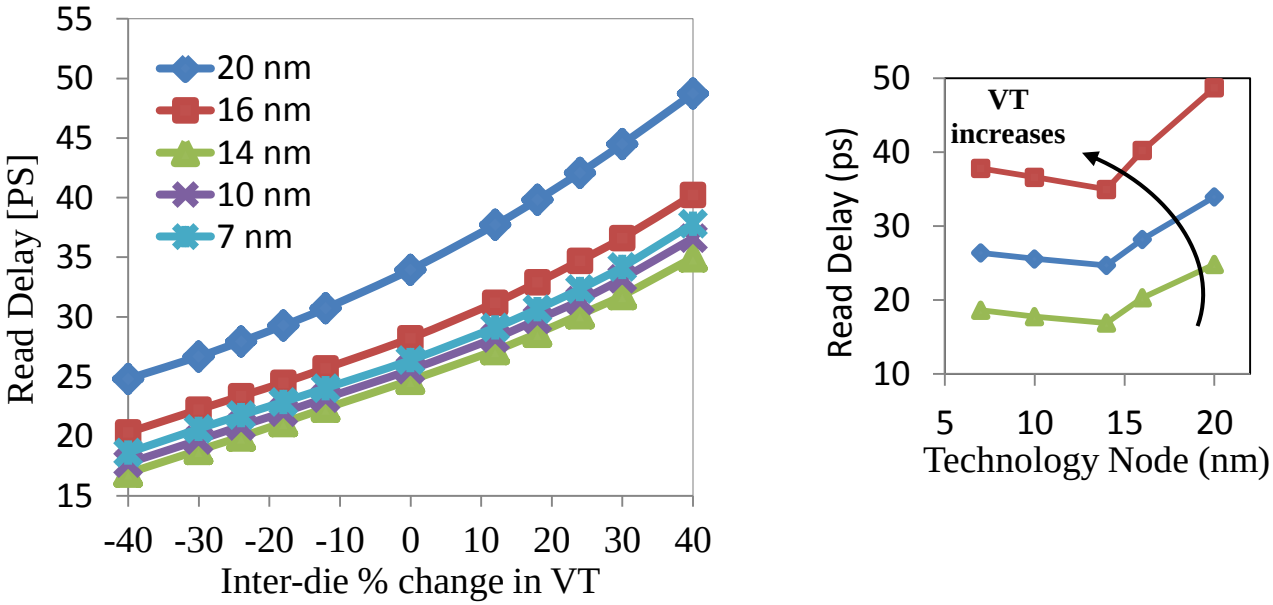


Figure 4-5: SRAM Read delay sensitivity to threshold voltage inter-die variations range of +/-40 % at various technology nodes from 20nm down to 7nm node.

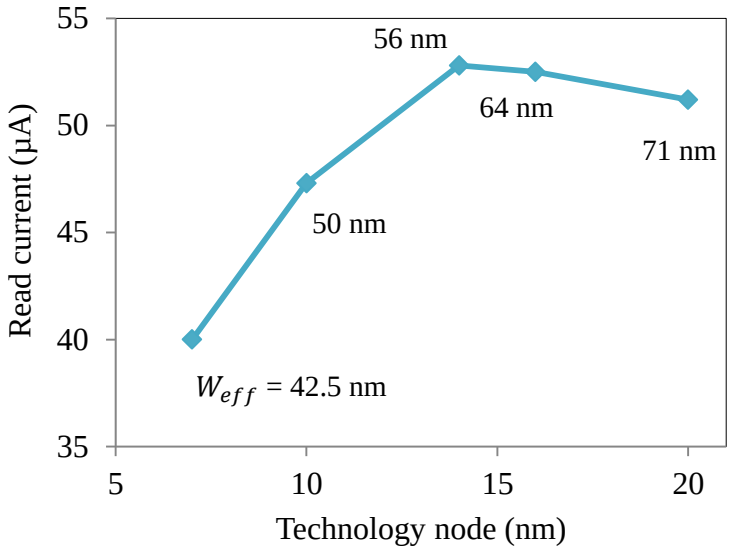


Figure 4-6: Device current per bit-cell with technology scaling from 20nm to 7nm node, where Weff = 2Hfin+Tfin, and Wtot = Nfin Weff.

This in fact has an adverse effect on the read current as it can be seen in Figure **4-6**; the current is almost constant (slightly increasing) as we go from the 20 nm to 16 nm and 14 nm nodes, however it drops at the 10 nm and further decreases reaching the 7 nm node.

So despite the fact that with technology scaling the current value per unit width is expected to increase, the current per bit-cell is decreasing as a result of the adopted scaling strategies to keep SCEs under control, since scaling both T_{fin}, and H_{fin} reduces the effective channel width.

Consequently, this raises a serious challenge for SRAM design in extremely scaled technology nodes, since this fact implies that to retrieve this loss of performance, keeping SCEs under control, some cell devices generally will need to be sized up which contradicts the trend of higher density SRAM arrays with scaling.

B. Power consumption

Power consumption is one of the critical metrics for any logic circuits and analyzing the scaling trends of both the active and leakage components is of special concern. From one hand, as the technology scales, all sources of leakage power increase. Shrinking the channel length increases the sub-threshold leakage component and scaling the oxide thickness severely affects the gate tunneling current which is another component of the total leakage current.

From the other hand, the increased variability sources with scaling and the resulting effect on the threshold voltage spread significantly impacts the leakage power due the exponential dependence on V_T. Figure **4-7** shows the percentage of the leakage power component to the active power component and its sensitivity with the threshold voltage variations at both 20nm and 7nm nodes. As it can be seen, for the 20nm node, as V_T decreases the amount of the leakage power increases and contributes significantly a larger portion of the total power consumption. While, for the 7nm, the leakage power already occupies a significant portion of the total power and changing the threshold voltage has a minor impact on the relative percentage.

It can be concluded that, as the technology scales, the leakage power component increases and occupies a significant portion of the total power consumption, however the variability of the leakage power to the V_T variation significantly reduces. This fact has further implications on other performance metrics as it will be discussed in the next section.

C. Static Noise Margins

Figure **4-8** shows the read and write static noise margins (RSNM, WSNM) with technology scaling. As it can be seen in Figure **4-8**.a) the RSNM shows the same behavior of the read delay with a degradation starting from the 10nm node to the 7nm.

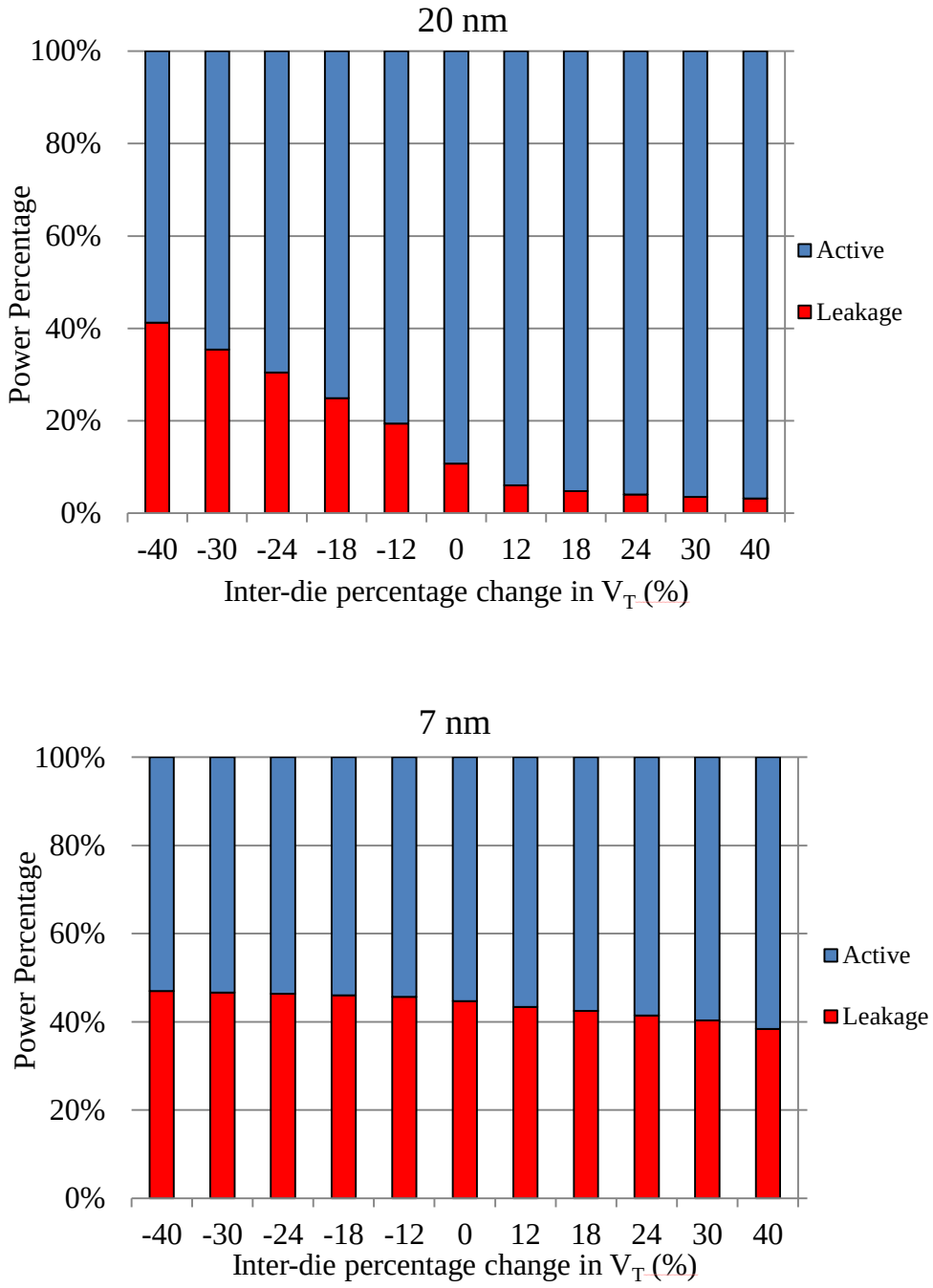


Figure 4-7: Sensitivity of the percentage leakage power to the active power with threshold voltage variations; a) 20nm node, b) 7nm node.

This also can be attributed to the degradation of the read current as discussed in the above section which affects the read operation as a whole from both the delay and stability point of view. For the WSNM, a clear degradation of 28% at 7nm with respect to its value at the 20nm node can be shown in Figure **4-8**.b), which is primarily as a result of scaling the supply voltage. Figure **4-9** shows the sensitivity of the RSNM and WSNM with the V_T variations at the 20nm and 7nm technology nodes. First it can be seen in Figure **4-9**.a) that with deceasing V_T the RSNM is degraded for both the technology nodes, since reducing V_T increases the leakage current which in turn increases the voltage of the node to be read (assuming read '0') leading to an increase in the probability of destructive read operation. In addition, reducing V_T affects the VTC of the inverters which affects the trip point making it easier for the '0' storage node to flip to '1'. Second, the degradation in the RSNM for the 20nm node is around +/- 25 % and for the 7nm is just about +/- 10 % with respect to the value at the nominal threshold voltage.

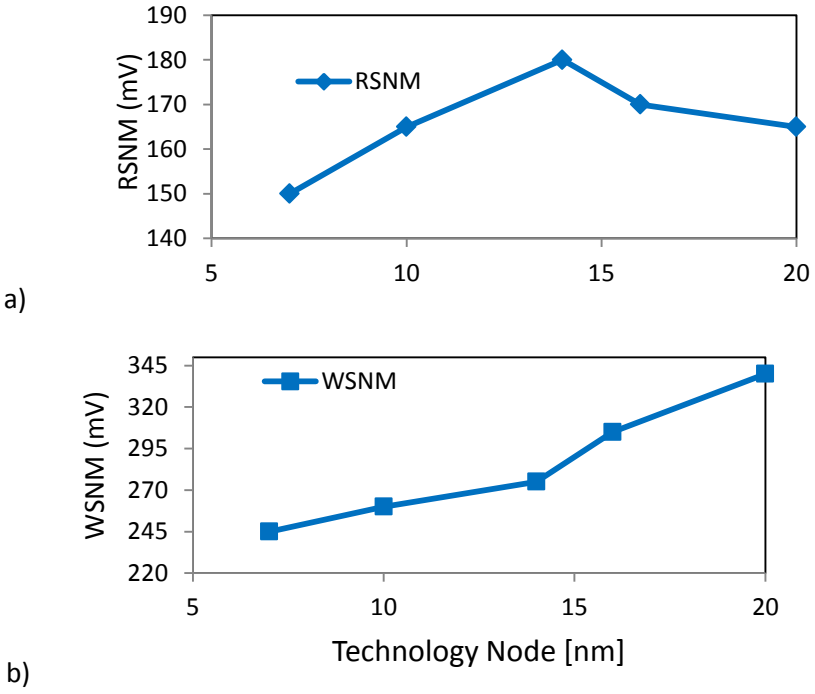
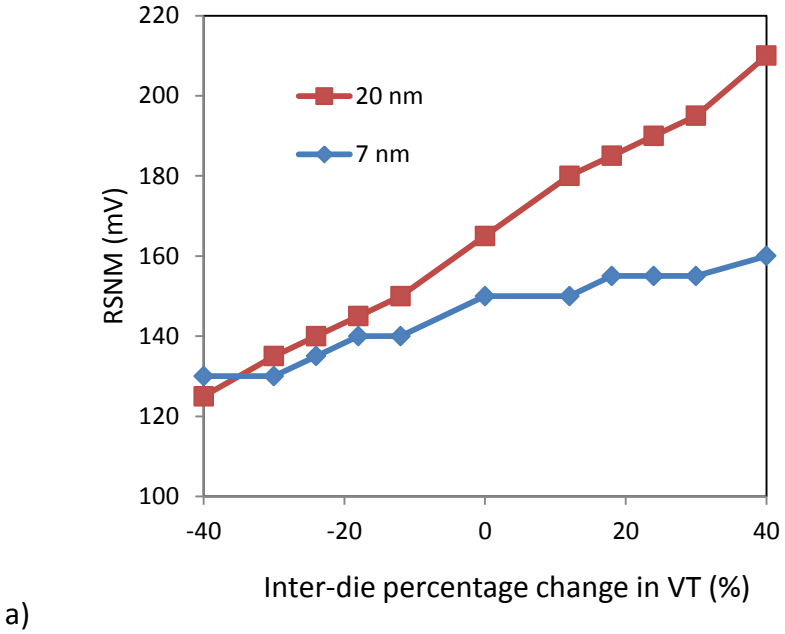
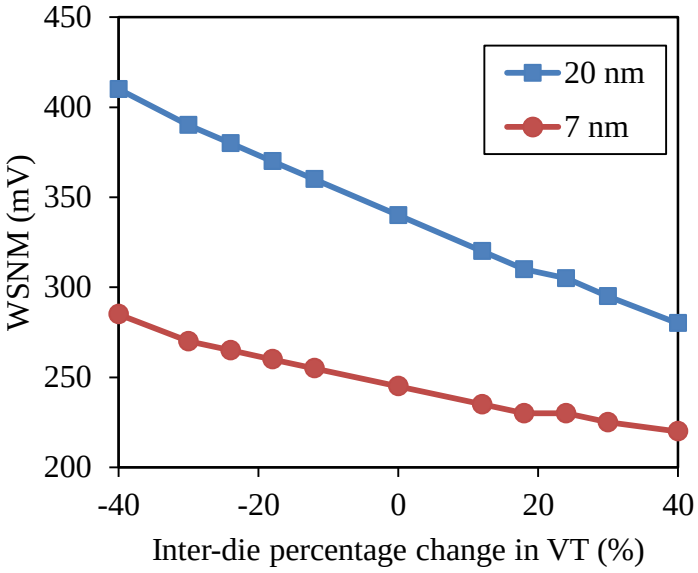


Figure 4-8: Read and write static noise margins with technology scaling



a)



b)

Figure 4-9: Sensitivity of the read and write noise margins to the threshold voltage variations for 20nm and 7nm technology nodes; a) RSNM, b) WSNM

This behavior can be explained as a result of less sensitivity of the leakage current to the V_T variations with technology scaling compared to that at the 20nm as discussed in the above section. Figure **4-9**.b) shows the sensitivity of the WSNM to the V_T variation showing the opposing response to the RSNM as it enhances with decreasing V_T. In addition the percentage change in WSNM for both the technologies is quite closer as compared to the RSNM.

4.2.3. Conclusion

The performance of FinFET 6T SRAM of 256-bit cell is evaluated with technology scaling. The impact of a given range of threshold voltage variations on basic performance metrics is reported. The results show that, starting from the 10nm node and down to the 7nm, clear performance degradation is observed in the read operation impacting both the delay and stability metrics.

The degradation of the read current per bit-cell with technology scaling as a result of scaling other parameters besides the channel length was seen to be the main reason behind the observed degradation in the read operation.

The study also shows that, with technology scaling, the leakage power occupies larger portion of the total power consumption, however the sensitivity of the leakage to threshold variations is reduced with scaling down the technology.

4.3. Analysis and Optimization for Dynamic Read Stability in 28nm SRAM Bit-cells

In this section we move to the second study in evaluating circuit design with nano-scale transistors. We keep using the SRAM unit as the main block but as we mentioned at the beginning of this chapter this is more advanced study where we investigate more sophisticated operation and design characteristics. This part is only for the 28nm FD-SOI device which we consider as a peer technology to the FinFET devices, since it the most advanced available commercial PDK.

According to the International Technology Roadmap for Semiconductors (ITRS-2013) [*66*], major challenge of SRAM scaling is to maintain adequate noise margins and control key instabilities and soft error rates in the presence of random threshold voltage (V_T) fluctuations. Static noise margin (SNM) has been used as a mainstream method for SRAM stability characterization [*67*].

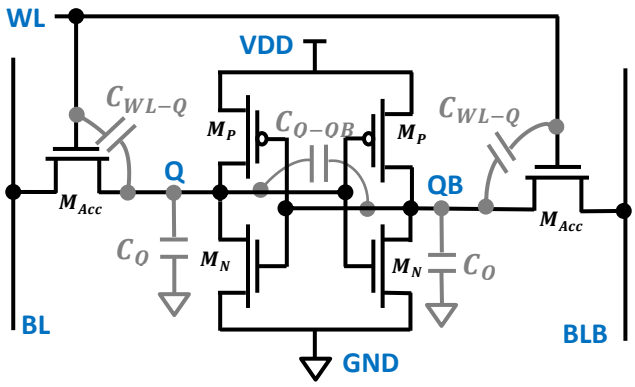


Figure 4-10: Conventional 6T SRAM cell with the main parasitic capacitances, under study, contributing to the dynamic effects.

It is based on evaluating the voltage transfer characteristics (VTC) of the cross-coupled inverters through a DC simulation or measurement holding both bitlines (BL, BLB) and wordline (WL) at the DC supply voltage (V_{DD}). Neglecting the dynamic effects such as the finite duration of the WL pulse width and the precharging of bitlines, SNM yields pessimistic read stability value as compared to the dynamic metrics, thereby imposing an extra burden in the design process [*68*]. With the continuous technology scaling, the design space is eventually narrowing down due to the associated increase in process variations and voltage scaling [*69*], [*70*].

As a result, it becomes more and more difficult to afford overestimated constraints. In [*71*], static read margins were observed to overestimate failures by 10-100 X. Moreover, since SNM metrics are static in nature and neglect any time-dependence since they are based on DC simulations, they do not consider dynamic effects such as charge sharing and capacitive coupling associated with the cell's parasitic capacitances.

Several works have addressed dynamic metrics for SRAM stability analysis [*71*], [*72*], [*73*], [*74*]. Some were introduced in terms of time metrics [*73*], others were kept as voltage metrics, but the bottom-line is that all are based on transient simulations that take the dynamic non-linear effects of the SRAM cell into account.

Others proposed semi-analytical models [*75*], based on simple approximated circuit equations in time domain. In [*73*], the DNM was analyzed based on the stability boundary theory (or Separatrix) and the possible correlation between DNM and SNM was examined. Practically, the importance of the dynamic analysis for SRAM operation increases as a result of increased timing constraints and development of new dynamic read/write assist techniques [*70*], [*76*], [*77*]. Nevertheless, a quantitative study for the basic dynamic effects and their different contributions to the difference between the DNM and SNM metrics has not been reported up to now.

In this study, we analyze the dynamic stability through the DNM and its dependence on different dynamic effects including the WL pulse width and the number of cells per column (bitline). In addition, we extend the work to study the effect of different parasitic capacitances within the 6T SRAM cell, shown in Figure 4-10, on the DNM. We present the evolution from SNM to DNM through cumulative dynamic effects showing the contribution of each effect, and define the parameters' limits for the convergence of DNM to SNM. Finally, we introduce a comparative example of bit cell sizing for SNM and DNM. The results have been obtained for the conventional 6T SRAM operating at 1V in 28 nm FDSOI CMOS.

4.3.1. Quantitative analysis of Dynamic Read Noise Margin

The limitation of the static read noise margins (SNM) comes from the fact that BL, BLB and WL are all driven to V_{DD} in the stability DC characterization setup. In transient operation and dynamic characterization setup, the story is different. Indeed, first in practice the fact that the WL is pulsed in transient operation means that the SRAM internal storage nodes are not infinitely susceptible to flip.

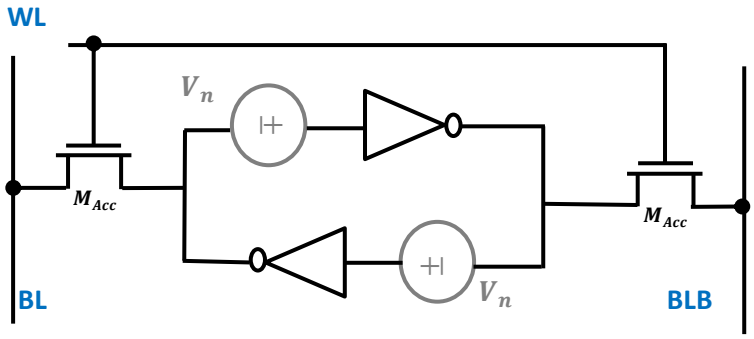


Figure 4-11: Equivalent circuit for DNM characterization setup.

Second, BL and BLB are not tied to V_{DD} but rather are initially pre-charged, then left floating and getting discharged by the cell, which limits the contention between the pull down device (M_N) in holding the '0' and the discharging current from the BL. As a result, the static analyses are known to yield pessimistic noise margins during the read operation.

The characterization setup for DNM is shown in Figure 4-11 which is considered in a dynamic manner by means of transient simulations. The evolution from SNM to DNM through cumulative dynamic effects is shown in Figure 4-12. By looking at the SRAM operation and its control signals, the main dynamic effects can be summarized as follows:

(A) WL opening:

The sudden change from '0' to '1' at the assertion of WL pulse has a critical impact on the cell's stability due to the capacitive coupling between WL signal at the gate of the access transistors (M_{Acc}) and the internal storage node (Q). The amount of the coupling depends on the ratio between the coupling capacitance between the two nodes to the self-capacitance of node Q. In some cases; such sudden changes might cause the node Q to flip leading to destructive read operation.

(B) Finite WL pulse width (PW):

This is also a characteristic for the WL signal. During this time the access transistors are ON, connecting the bitlines to the internal storage nodes.

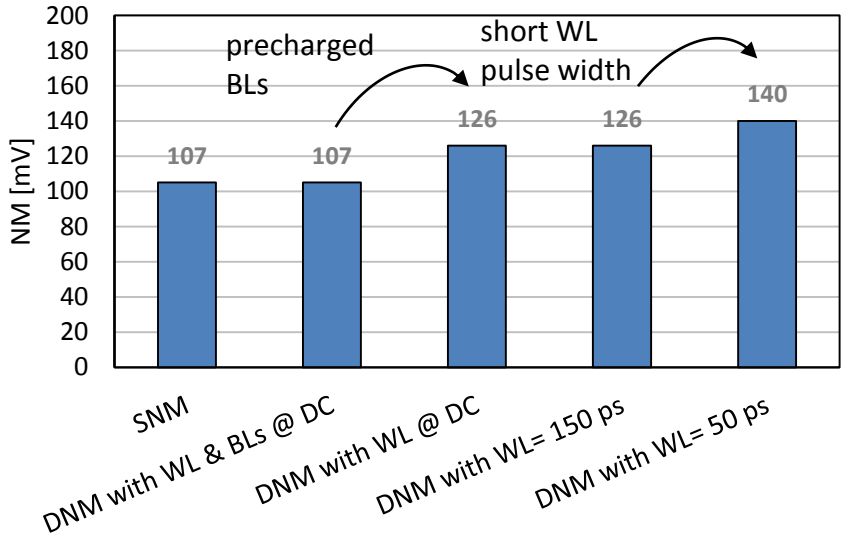


Figure 4-12: Evolution of noise margin from SNM to DNM with cumulative dynamic effects The BL discharge time for 100mV of differential voltage is 22ps, which allows sufficient margin for a 50-ps WL pulse.

Apparently, the longer this time the worse for the read stability since it gives more opportunity for the node to flip its stored data. It is interesting to note that, as shown in Figure 4-13.a), there is a certain pulse width beyond which the DNM saturates to a certain limit.

On the other side, as we shrink the WL pulse width, the cell essentially enjoys larger DNM however the required time for proper read operation imposes a constraint on such shrinking. Thereby, for a very narrow WL pulses beyond the read delay limit, the cell might not be functional, hence results in a maximum achievable DNM.

(C) Discharging bitlines:

Precharging both BL and BLB to V_{DD} and then leaving them floating until the assertion of the WL to discharge through active M_{ACC} is a pure dynamic behavior that cannot be captured in DC simulation.

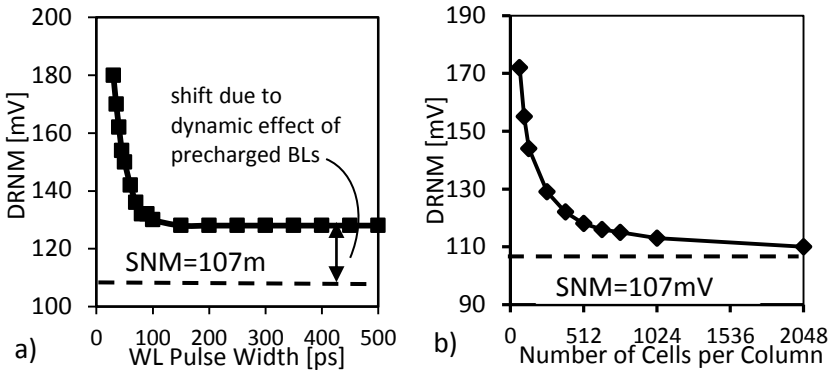


Figure 4-13: Behavior of DNM with a) increasing WL pulse width and b) increasing density of SRAM cell array

For further understanding, we examine this effect in the extreme limits by increasing the number of cells/bitline which is equivalent to increasing the bit-line capacitance C_{BL}. As it is shown in Figure 4-13.b), increasing the number of cells degrades the DNM and eventually converges to the SNM limit and becoming very close at 2048 cells/column. It should also be noted that this dynamic behavior is independent of the WL PW while it depends on the cell parasitic capacitances. Therefore, as shown in Figure 4-13.a), even at quite long WL PW, the DNM saturates at a certain level leaving a constant shift above the SNM.

(D) Frequency of cell access:

Another factor that takes the dynamic behavior into account is the time between successive read operations. In fact, two opposing effects happen with increasing the frequency. First, the time between successive operations reduces which means that another disturbance might come in before the node has fully recovered from the first disturbance of the first read, hence it will be much vulnerable to flip. Second, assuming that a reduction in the active access time (WL pulse width) is associated with increasing the frequency, as discussed in the above subsections, this results-in more reliable operation and higher stability hence makes this effect less critical.

4.3.2. Effect of parasitic capacitances on R/W dynamic noise margin:

The effect of parasitic capacitances on the behavior of DNM is worth to consider, since these capacitances characterize the dynamic behavior of the cell and essentially form the difference between the static and dynamic stability analyses. The considered components in this study, as shown in Figure **1-2**, are: (i) self-capacitance of the storage nodes (C_Q); (ii) coupling capacitance between the storage nodes (C_{Q-QB}); and (iii) coupling capacitance between the WL signal and the storage nodes (C_{WL-Q}), in addition to (iv) the bitline capacitance (C_{BL}) whose effect is implicitly captured when varying the number of cells per column in the previous section (Figure 4-13).

First, adding capacitance at the storage nodes improves the cell stability as shown in Figure 4-14, since the charge sharing increases and distributes the charges better at the storage nodes which helps maintaining the stored data at Q and QB and consequently results in improved stability levels. Second, it was found that the stability improvement is different from one component to the other. The transient waveforms of Q and QB are shown in Figure 4-15, where the noise source (V_n) is swept through a successive transient simulations with a step of 10mV until the cell flips its data, for each capacitance component. It can be seen that, the DNM for 2fF C_{Q-QB} is about 190mV instead of 170mV for 2fF C_Q at relatively long WL pulse width (500ps). This can be attributed to the enhanced Miller effect linked to this capacitance, since the feedback loop within the cross coupled inverters works on compensating any disturbance during the read operation yielding better stability levels. It can be also seen that a 2fF C_{WL-Q}yields even better DNM of 210mV. This can be explained as a result of the voltage overshoot due to the coupling effects which is quite high in the case of C_{WL-Q} due to the direct coupling with the WL signal. Although such coupling effects might seem to degrade the overall stability, due to the differential nature of the 6T bitcell, the overshoot in the QB node enhances the drive strength of the pull-down NMOS in holding the '0' at Q node, hence yields improved DNM levels. Similar results for DWNM are shown in the right column.

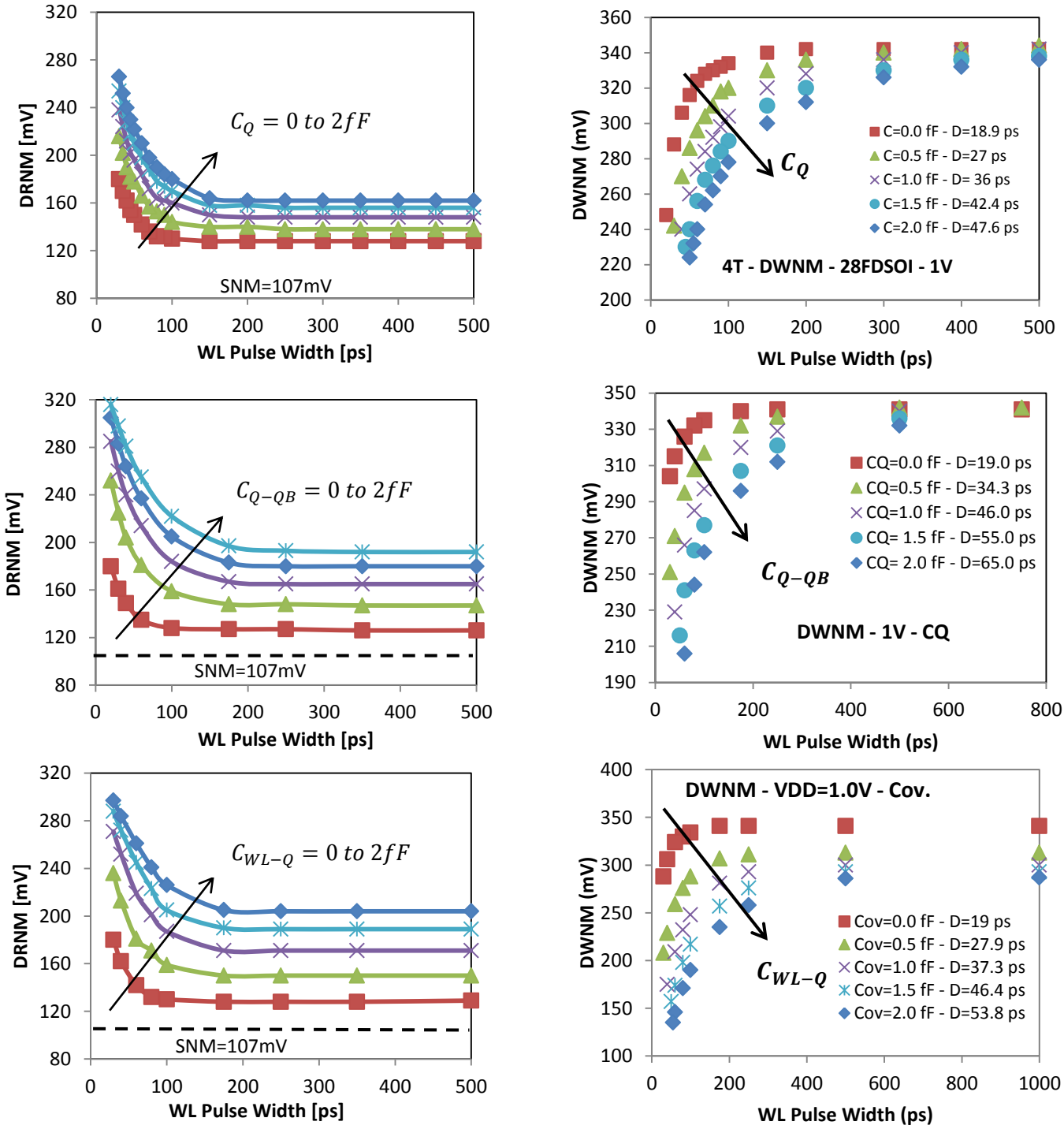


Figure 4-14: The behavior of DNM with changing the pulse width of the WL signal and varying the different parasitic capacitance components from 0 to 2fF: a) self-capacitance of the storage nodes (C_Q); b) coupling capacitance between the storage nodes ($C_{(Q-QB)}$); and c) coupling capacitance between the WL signal and the storage nodes ($\boldsymbol{C_{WL-Q}}$).

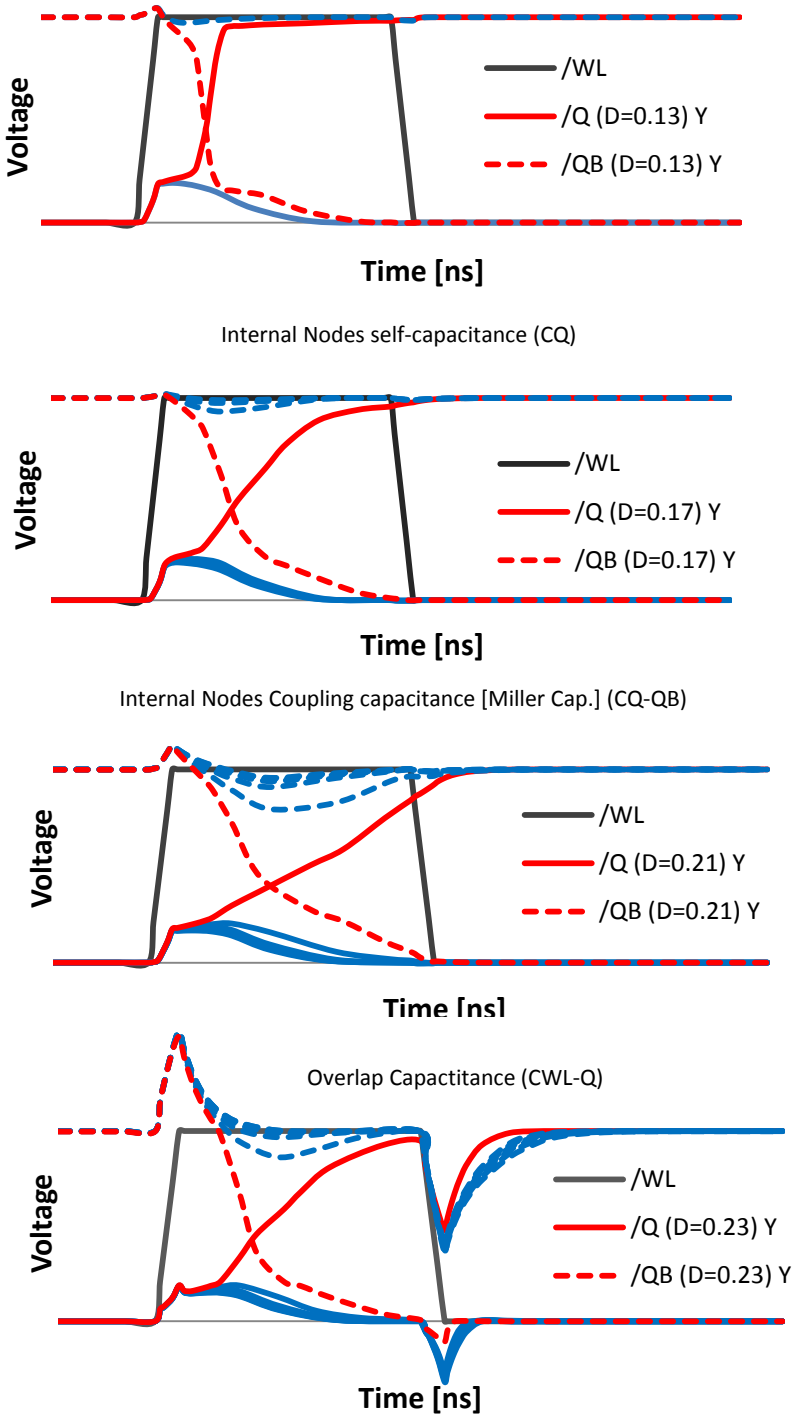


Figure 4-15: Transient waveforms for Q, QB at different noise levels (Vn) to the level at which it flips its data, along with WL signal, a) intrinsic case (cell parasitic capacitances at the used sizing), b) added CQ, c) added CQ-QB, d) added CWL-Q.

4.3.3. Sizing for DNM: Design Perspective

Let us now have a look on how the consideration of read DNM instead of read SNM allows downsizing the β ratio of the cell (width ratio between the pull-down NMOS and the access transistors) under a fixed noise margin constraint. Figure **4-16** shows that for a wide range of β ratio, the DNM is 40mV higher than the SNM, thereby resulting in a potential 40% reduction of the β ratio to reach a 150-mV noise margin.

Previous works proposed to intentionally add extra capacitors to improve the performance either using the gate capacitance of MOS transistor [*78*], or a fringe metal capacitor placed above the six transistors of the cell as in [*79*]. Based on Section III, we could drastically improve the DNM by adding extra impact of C_{WL-Q}capacitance.

Figure **4-16** shows that the addition of two 0.5fF further significantly improves the DNM. For example, for 150mV SNM, 2.1 β ratio is required, while in the case of design for DNM with the proposed modification, only 0.6 β ratio is sufficient to achieve same level of stability. Such a strong reduction of the β ratio can be useful to save area and leakage (downsizing the NMOS pull down transistors) as well as to improve the writeability of the cell.

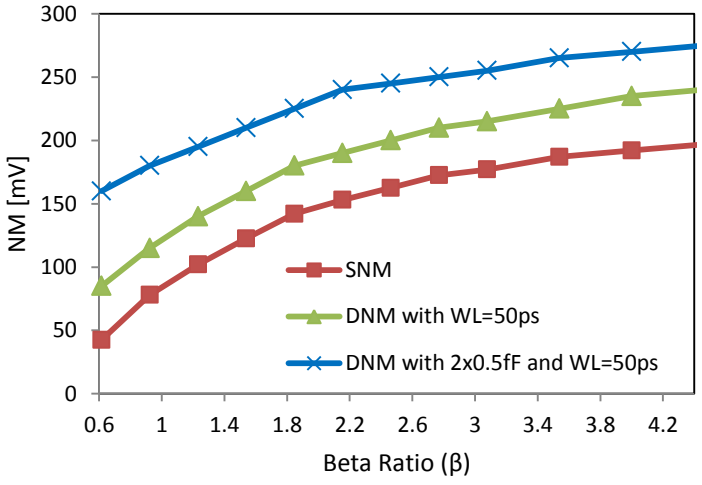


Figure 4-16: Dependence of read noise margins on the beta ratio under different conditions

4.3.4. Conclusions

The dynamic read noise margin (DNM) is quantitatively analyzed in in 28nm FDSOI CMOS based on cumulative dynamic effects through transient simulations showing the contribution of each effect. It was found that DNM is improved compared to SNM mainly because of the finite WL pulse width and the BL discharge. We reported that extending the pulse width beyond a certain limit (150ps in this case study) removes the dependence on the WL signal and cancels this first improvement. In addition, increasing the cell density to about 2048 cells/column further limits the second one. Consequently a transient characterization setup under such conditions yields the same noise margin from pure DC simulation.

The impact of different parasitic capacitances on the DNM behavior was discussed showing the respective stability improvement of increasing those attached to the storage nodes, and in contrary the degradation of increasing the bitline capacitance through higher density arrays.

From a design sizing perspective, the dependence of SNM, DNM, DNM with an extra 0.5fF C_{WL-Q} on the beta ratio was compared. It was found that the addition of 0.5fF C_{WL-Q} results in significant DNM improvement thereby allowing a 3.5× reduction of the β ratio while keeping a 150-mV DNM (compared to sizing for a 150-mV SNM).

5. CONCLUSIONS AND OUTLOOK

5.1. Summary

In Chapter 2, we discussed most of the transport models used in TCAD for simulating nano-scale transistors. We benchmarked three different versions of the drift-diffusion model against Monte Carlo model in a trial to come up with more computationally efficient model that can be used for simulating nano-scale TG FinFET. After analyzing the results, we can conclude that conventional models, such as the DD, already use enough number of coefficients and fitting parameters in modeling different physical quantities, such as the mobility and velocity models. The use of fitting parameters enables matching broad range of experiments and characterization curves obtained using sophisticated models such as Monte Carlo, however at a specific set of conditions (channel length, bias, structure, ..). Therefore, being able to come up with a universal model that takes all these different parameters and conditions into account is difficult if not impossible in addition to being quite non-physical, since these approaches rely on fitting techniques in the first place. Consequently, even if we got a match using a specific set of parameters with a specific characteristic curve, this doesn't guarantee in any way to have same match with other characteristics or quantities that we may not have access to assess its validity or most probably we need to study.

In Chapter 3, based on the conclusion from the previous chapter, we considered Monte Carlo techniques in all our simulations, despite being computationally extensive (it might take more than one day to get one IV characteristic). A numerical study for TG FinFET was carried out to study the scaling behavior of such devices in the light of the most recent international technology roadmap for semiconductors. Monte Carlo models are adjusted to simulate both ballistic (by switching off all the scattering mechanisms) and quasi-ballistic (what MC normally considers, including all the possible scattering

mechanisms) regimes for each channel length under consideration. The ballisticity ratio was extracted and discussed highlighting other reported values in the literature.

It was found that initially BR increases slowly with shrinking L, during this scaling range, the high field region is quite far from the ToB. With further scaling, the high field region gets closer and closer to the ToB. Once the difference between the ToB and the inflection point on the CB profile is around 8.5 KT or less, the BR is significantly impacted exhibiting a sudden increase before it slows down again approaching 90% when eventually the two regions merge and points X, Y coincide.

Despite the reported improvement in the ballisticity with scaling, the devices are not able to attain expected performance improvements. This can be seen in (i) the diminishing improvement of the on-current with scaling the channel length of TG-FinFET, (ii) the increasing SCEs specially starting from 11.6 nm channel length despite the adopted scaling strategy.

The velocity profiles, for both the forward and backward components, along the channel were analyzed at different channel lengths. The simulation results reveal that the electron transport along the channel can be characterized by two critical points dividing the velocity/CB profiles into three spatial regions with significant change in the electron acceleration/energy. However, as the gate length shrinks the two points reduce to a single point; hence the velocity profile is characterized by two regions instead of three as a result of the dominance of the off-equilibrium phenomena with extreme length scaling.

In Chapter 4, we extended the study to include the circuit design level. First, the performance of FinFET 6T SRAM of 256-bit cell is evaluated with technology scaling. The impact of a given range of threshold voltage variations on basic performance metrics is reported. The results show that, starting from the 10nm node and down to the 7nm, clear performance degradation is observed in the read operation impacting both the delay and stability metrics.

The degradation of the read current per bit-cell with technology scaling as a result of scaling other parameters besides the channel length was seen to be the main reason behind the observed degradation in the read operation.

The study also shows that, with technology scaling, the leakage power occupies larger portion of the total power consumption, however the sensitivity of the leakage to threshold variations is reduced with scaling down the technology.

In the second part, we switched to other peer technology (28m FDSOI) for the purpose of carrying out a comparative study. The dynamic read noise margin (DNM) is quantitatively analyzed in in 28nm FDSOI CMOS based on cumulative dynamic effects through transient simulations showing the contribution of each effect. It was found that DNM is improved compared to SNM mainly because of the finite WL pulse width and the BL discharge. We reported that extending the pulse width beyond a certain limit (150ps in this case study) removes the dependence on the WL signal and cancels this first improvement. In addition, increasing the cell density to about 2048 cells/column further limits the second one. Consequently a transient characterization setup under such conditions yields the same noise margin from pure DC simulation.

The impact of different parasitic capacitances on the DNM behavior was discussed showing the respective stability improvement of increasing those attached to the storage nodes, and in contrary the degradation of increasing the bitline capacitance through higher density arrays.

From a design sizing perspective, the dependence of SNM, DNM, DNM with an extra 0.5fF C_{WL-Q} on the beta ratio was compared. It was found that the addition of 0.5fF C_{WL-Q} results in significant DNM improvement thereby allowing a 3.5× reduction of the β ratio while keeping a 150-mV DNM (compared to sizing for a 150-mV SNM).

5.2. Outlook

5.2.1. On the device level

Based on the conclusions deduced in Chapter 3, in addition to performing more simulations at different bias conditions for the electron average velocity and its energy along the channel, compact models can be developed to describe the electron transport in nana-scale devices working near ballistic and be more physics-based trying to reduce the huge amount of fitting parameter to match the performance of actual devices in fast circuit simulations.

5.2.2. On the circuit level

In the light of the Monte Carlo simulations performed in 3-D, optimization for the predictive technology models (PTM) is required to improve its predictability for the performance especially for the extremely scaled technology nodes.

Repeating same study, discussed in second part of Chapter.4, about the stability of SRAM cells and the evolution from dynamic to static metric in addition to investigating the design example for dynamic metrics instead of static ones, using TG FinFET to assess its potential benefits over most advanced peer technology and how the technology scaling impacts such benefits.

REFERENCES

[1] L. Witters et al, "Strained Germanium quantum well pMOS FinFETs fabricated on in situ phosphorus-doped SiGe strain relaxed buffer layers using a replacement Fin process," in *in Proc. IEDM*, 2013.

[2] K. S. Novoselov et al, "A roadmap for graphene," *Nature*, vol. 490, pp. 192-200, 2012.

[3] K., Cavin, R. K., Porod, W., Seabaugh A. C. & Welser, J. Bernstein, "Device and Architecture Outlook for Beyond CMOS Switches," *Proc. IEEE*, vol. 98, pp. 2169–2184 , 2010.

[4] "Room temperature superfluidity in graphene bilayers," *Phys. Rev. B*, vol. 78, 2008.

[5] Pojen Chuang et al, "All-electric all-semiconductor spin field-effect transistors," *Natur Nanotechnology*, vol. 10, pp. 35-39, 2015.

[6] J. P. Colinge, "Multiple-gate SOI MOSFETs," *Solid State Electron*, vol. 48, pp. 897–905 , 2004.

[7] Isabelle Ferain et al, "Multigate transistors as the future of classical metal–oxide–semiconductor field-effect transistors," *Nature* , vol. 479, pp. 310-316, 2010.

[8] J. G. Fossum, L. Mathew, M. M. Chowdhury, W. Zhang, G. O. Workman, and B.-Y. Nguyen V. P. Trivedi, "Physics-based compact modeling for nonclassical CMOS," in *IEEE/ACM International Conference on Computer-Aided Design, ICCAD-2005.* , 2005, pp. 211-216.

[9] S.Y. Chou, D. A. Antoniadis, and H. I. Smith, "Observation of electron velocity overshoot in sub-100-nm-channel MOSFET's in Si," *IEEE Electron Device Lett.*, pp. 665-667, 1985.

[10] G. Timp, J. Bude, K.K. Bourdelle, J. Garno, A. Ghetti, H. Gossmann, M. Green, G. Forsyth, Y. Kim, R. Kleiman, "The Ballistic Nano-transistor," in *IEDM Tech. Dig.*, 1999, pp. 55-58.

[11] Mark Lundstrom, "Elementary Scattering Theory of the Si MOSFET," *IEEE Electron Device Lett.*, vol. 18, no. 7, pp. 361-363, July 1997.

[12] Dragica Vasileska and Stephen M. Goodnick, *Computational Electronics*.: Morgan & Claypool, 2006.

[13] M. Shur, *Physics of Semiconductor Devices*. New Jersey: Prentice Hall Series in Solid State Physical, 1990.

[14] D. L. Scharfetter and D. L. Gummel, "Large signal analysis of a Silicon Read diode," *IEEE*

Trans. Electron. Devices, vol. 16, pp. 64-77, 1969.

[15] Robert F. Pierret, *Semiconductor Device Fundamentals*.: Addison-Wesley Publishing Cempany, Inc., 1996.

[16] J.D. Bude, "MOSFET Modeling Into the Ballistic Regime," in *Simulation of Semiconductor Processes and Devices, 2000. SISPAD 2000. 2000 International Conference on*, 2000, pp. 23-26.

[17] Jung-Hoon Rhew, "PHYSICS AND SIMULATION OF QUASI-BALLISTIC TRA N SPO RT IN NANOSCALE TRANSISTORS," Purdue University , Thesis 2003.

[18] V. M. Polyakov, F. Schwierz, M. Kittler and T. Doll R. Granzner, "On the suitability of DD and HD models for the simulation of nanometer double-gate MOSFETs," *Phys. E, Low-Dimensional Syst. Nanostruct.*, vol. 19, no. 2, pp. 33-38, 2003.

[19] Synopsys, "Sentaurus Device User Manual," Synopsys, 2013.

[20] H. Tian, R. J. Trew, M. A. Littlejohn, and K. W. Kim D. L. Woolard, "Hydrodynamic electron-transport model: Nonparabolic corrections to the streaming terms," *Phys. Rev. B*, vol. 44, no. 20, pp. 11 119–11 132., 1991.

[21] Kausar Banoo et al, "Simulating quasi-ballistic transport in Si nanotransistors," *7th International Workshop on Comp. Elec.*, pp. 8-9, 2000.

[22] Synopsys Technical Notes, "Three-dimensional Device Simulations of 10 nm FinFETs Using Monte Carlo Model and Drift-Diffusion Model With Ballistic Mobility," 2013.

[23] Mark S. Lundstrom, Fellow, IEEE, and Dimitri A. Antoniadis, Fellow, IEEE, "Compact Models and the Physics of Nanoscale FETs," *IEEE TRANSACTIONS ON ELECTRON DEVICES*, vol. 61, FEBRUARY 2014.

[24] Mark S. Lundstrom, "Fundamentals of carrier transport," *Cambridge University press*.

[25] M. Nedjalkov, S. Selberherr H. Kosina, "The Stationary Monte Carlo Method for Device Simulation. I. Theory," *Journal of Applied Physics*, 2003.

[26] V.M. Polyakov, F. Schwierz, M. Kittler, T. Doll Ralf Granzner, "On the suitability of DD and HD models for the simulation of nanometer double-gate MOSFETs," *Science direct, Physica E 19 33 – 38*, 2003.

[27] Mark S. Lundstrom, "On the Mobility Versus Drain Current Relation for a Nanoscale MOSFET," *Trans. Electron Devices*, vol. 22, no. 6, pp. 293-295, 2001.

[28] M. Zilli et al, "On the Apparent Mobility in Nanometric n-MOSFETs," *IEEE Electron Device Letters*, vol. 28, no. 11, pp. 1036–1039, 2007.

[29] S. Y. Chou, D. A. Antoniadis, and H. I. Smith, "Observation of Electron Velocity Overshoot in Sub- 100-nm-channel MOSFET's in Silicon," *IEEE Electron Device Lett.*, vol. EDL-6, pp. 665-667, 1985.

[30] Mark Lundstrom, and Zhibin Ren, "Essential Physics of Carrier Transport in Nanoscale MOSFETs," *IEEE Trans. Electron Devices*, vol. 49, no. 1, pp. 133-141, Jan. 2002.

[31] B. Winstead and U. Ravaioli, "A quantum correction based on Schrodinger equation applied to Monte Carlo device simulation," *IEEE Trans. Electron Devices*, vol. 50, no. 12, pp. 2467–2473, Dec. 2003.

[32] R. Ravishankar, G. Kathawala, and U. Ravaioli, S. Hasan, and M. Lundstrom, "Comparison of Monte Carlo and NEGF Simulations of Double Gate MOSFETs," *J. Computational Electronics*, pp. 39–43, 2005.

[33] C. J acoboni, and P. Lugli, *The Monte Carlo method of semiconductor device simulation*.: Springer, 1989.

[34] Mark S. Lundstrom, *Fundamentals of carrier transport*.: Cambridge University press, 2000.

[35] D. J. Frank, S. E. Laux, and M. V. Fischetti, "Monte Carlo simulation of a 30 nm dual-gate MOSFET: How short can Si go?," in *IEDM Tech. Dig.*, 1992, pp. 553-556.

[36] M. Mouis, and S. Barraud , "Velocity distribution of electrons along the channel of nanoscale MOS transistors," in *Proc. ESSDERC*, 2003, pp. 147-150.

[37] Pierpaolo Palestri, David Esseni, Simone Eminente, Claudio Fiegna, Enrico Sangiorgi, and Luca Selmi., "Understanding Quasi-Ballistic Transport in Nano-MOSFETs: Part I—Scattering in the Channel and in the Drain," *IEEE Trans. Electron Devices.*, vol. 52, no. 12, pp. 2727-2735, Dec. 2005.

[38] Simone Eminente, David Esseni, Pierpaolo Palestri, Claudio Fiegna, Luca Selmi, and Enrico Sangiorgi, "Understanding Quasi-Ballistic Transport in Nano-MOSFETs: Part II—Technology Scaling Along the ITRS," *IEEE Trans. Electron Devices*, vol. 52, no. 12, pp. 2736-2743, Dec.

2005.

[39] E. Fuchs, P. Dollfus, G. L. Carval, S. Barraud, D. Villanueva, F. Salvetti, "A New Backscattering Model Giving a Description of the Quasi-Ballistic Transport in Nano-MOSFET," *IEEE Trans. Electron Devices*, vol. 52, no. 10, pp. 2280-2289, Oct. 2005.

[40] Jérôme Saint Martin, Arnaud Bournel, and Philippe Dollfus, "On the Ballistic Transport in Nanometer-Scaled DG MOSFETs," *IEEE Trans. Electron Devices*, vol. 51, no. 7, pp. 1148-1155, July 2004.

[41] Hideaki Tsuchiya, Kazuya Fujii, Takashi Mori, and Tanroku Miyoshi., "A Quantum-Corrected Monte Carlo Study on Quasi-Ballistic Transport in Nanoscale MOSFETs," *IEEE Trans. Electron Devices*, vol. 53, no. 12, pp. 2965-2971, Dec. 2006.

[42] Massimo V. Fischetti, Terrance P. O'Regan, Sudarshan Narayanan, Catherine Sachs, Seonghoon Jin, Jiseok Kim, and Yan Zhang., "Theoretical Study of Some Physical Aspects of Electronic Transport in nMOSFETs at the 10-nm Gate-Length," *IEEE Trans. Electron Devices*, vol. 54, no. 9, pp. 2116-2136, Sep. 2007.

[43] M. V. Fischetti, L. Wang, B. Yu, C. Sachs, P. M. Asbeck, Y. Taur, et al., "Simulation of electron transport in high-mobility MOSFETs: Density of states bottleneck and source starvation," in *IEDM Tech. Dig.*, Dec., 2007, pp. 109-112.

[44] Hamdy Abd El Hamid, Jaume Roig Guitart, Valeria Kilchytska, Denis Flandre, and Benjamin Iñiguez, "A 3-D Analytical Physically Based Model for the Subthreshold Swing in Undoped Trigate FinFETs," *IEEE Trans. Electron Devices*, vol. 54, no. 9, pp. 2487-2496, Sep. 2007.

[45] HA Hamid, B Iñíguez, D Jiménez, LF Marsal, and J Pallarès, "A simple model of the nanoscale double gate MOSFET based on the flux method," *physica status solidi* , vol. 2, no. 8, pp. 3086-3089, May 2005.

[46] Manuel Aldegunde, Antonio Jesus García-Loureiro, and Karol Kalna, "3D Finite Element Monte Carlo Simulations of Multigate Nanoscale Transistors," *IEEE Trans. Electron Devices*, vol. 60, no. 5, pp. 1561-1567, May 2013.

[47] F.M. Bufler · L. Smith, "3D Monte Carlo simulation of FinFET and FDSOI devices with accurate quantum correction," *J. Computational Electronics*, vol. 10, pp. 651–657, 2013.

[48] Synopsys, Device Monte Carlo User's Guide, 2013.

[49] Yang Liu, Mathieu Luisier, Amlan Majumdar, Dimitri A. Antoniadis, and Mark S. Lundstrom, "On the Interpretation of Ballistic Injection Velocity in Deeply Scaled MOSFETs," *IEEE Trans. Electron Devices*, vol. 59, no. 4, pp. 994-1001, Apr. 2012.

[50] Ahmed T. El-Thakeb et al, "Performance evaluation of FinFET based SRAM under statistical VT variability," in *26th International Conference on Microelectronics (ICM)*, 2014, pp. 88-91.

[51] J Zhuge, R Huang RWang, "An Experimental Study on Carrier Transport in Silicon Nanowire Transistors: How Close to the Ballistic Limit?," in *in Solid-State and Integrated-Circuit Technology, ICSICT*, 2008, pp. 46-49.

[52] Anthony Lochtefeld, Dimitri A. Antoniadis, "On Experimental Determination of Carrier Velocity in Deeply Scaled NMOS: How Close to the Thermal Limit?," *IEEE Electron Device Lett*, vol. 22, no. 2, pp. 95-97, Feb. 2001.

[53] T Poiroux, J S-Martin, D Munteanu, J. Autran V. Barral, "Experimental Investigation on the Quasi-Ballistic Transport: Part I—Determination of a New Backs-cattering Coefficient Extraction Methodology," *IEEE Trans. Electron Devices*, vol. 56, no. 3, pp. 408-419, 2009.

[54] H Huan, K-Chuan M.-J. Chen, "Temperature dependent channel backscattering coefficients in nanoscale MOSFETs," in *in IEDM Tech. Dig.*, 2002, pp. 39-42.

[55] P Palestri, D Esseni M. Zilli, "On the experimental determination of channel back-scattering in nanoMOSFETs," in *in IEDM Tech. Dig.*, 2005, pp. 105-108.

[56] A Gnudi, S Reggiani E. Gnani, "Ballistic Ratio and Backscattering Coefficient in Short-Channel NW-FETs," in *in Eur. Solid-State Device Res. Conf., ESSDERC '09*, 2009, pp. 476- 479.

[57] M. S. Lundstrom, "A Landauer approach to nanoscale MOSFETs," *J. Computat. Electron.*, vol. 1, pp. 481-489, 2002.

[58] Robert F. Pierret, *Semiconductor Device Fundamentals*.: Addison-Wesley, 1996.

[59] Mark S. Lundstrom, Dimitri A. Antoniadis, "Compact Models and the Physics of Nanoscale FETs," *IEEE Trans. Electron Devices*, vol. 61, no. 2, pp. 225-233, Feb. 2014.

[60] Predictive Technology Model (PTM). http://ptm.asu.edu/.

[61] International Technology Roadmap of Semiconductors (ITRS). (2013) http://www.itrs.net/Links/2013ITRS/Home2013.htm.

[62] Seid Hadi Rasouli et al, "Design Optimization of FinFET Domino Logic Considering the Width Quantization Property," *IEEE Trans. Electron Devices*, vol. 57, no. 11, pp. 2934-2943, 2010.

[63] Mingu Kang et al, "FinFET SRAM Optimization With Fin Thickness and Surface Orientation," *IEEE Trans. Electron Devices*, vol. 57, no. 11, pp. 2785-2793, 2010.

[64] A. Asenov et al, "Simulation Based Transistor-SRAM Co-Design in the Presence of Statistical Variability and Reliability," in *in Tech. Dig. of IEDM*, 2013, pp. 33.1.1-33.1.4.

[65] Xingsheng Wang et al, "Impact of Statistical Variability and Charge Trapping on 14 nm SOI FinFET SRAM Cell Stability," in *in ESSDERC*, 2013, pp. 234-237.

[66] "International Technology Roadmap of Semiconductors ," http://www.itrs.net/Links/2013ITRS/Home2013.htm, 2013.

[67] K. Agarwal and S. R. Nassif, "Statistical analysis of sram cell stability," in *IEEE/ACM DAC*, 2006, pp. 57-62.

[68] M. Khellah, D. Khalil, D. Somasekhar, Y. Ismail, T. Karnik, and V. De, "Effect of power supply noise on SRAM dynamic stability," *Symp. VLSI Circuits.*, pp. 76–77, 2007.

[69] K. Roy S. Mukhopadhyay, H. Mahmoodi, "Modeling of failure probability and statistical design of SRAM array for yield enhancement in nanoscaled cmos.," *IEEE Trans. CAD*, vol. 24, no. 12, pp. 1859–1880, Dec. 2005.

[70] M. Khellah, Y. Ye, N. S. Kim, D. Somasekhar, G. Pandya,A. Farhang, K. Zhang, C. Webb, and V. De., "Wordline & bitline pulsing schemes for improving SRAM cell stability in low-vcc 65nm CMOS designs.," in *Symp. on VLSI Circuits*, June 2006, pp. 109-118.

[71] S. O. Toh, Z. Guo, T.-J. K. Liu, and B. Nikolic, "Characterization of Dynamic SRAM Stability in 45 nm CMOS," *IEEE J. Solid-State Circuits*, vol. 46, no. 11, pp. 2702-2712, Nov. 2011.

[72] D. Khalil et al, "Accurate estimation of SRAM dynamic stability," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 16, no. 12, pp. 1639–1647, Dec. 2008.

[73] W. Dong, P. Li, and G. Huang, "SRAM dynamic stability: Theory, variability and analysis," in *IEEE/ACM Internat. Conf. Computer-Aided Design, ICCAD 2008*, 2008, pp. 378–385.

[74] G. M. Huang, W. Dong, Y. Ho, and P. Li, "Tracing SRAM separatrix for dynamic noise margin analysis under device mismatch.," in *IEEE Int. Behavioral Modeling and Simulation Conf.*,

2007.

[75] B. Zhang, A. Arapostathis, S. Nassif, and M. Orshansky, "Analytical modeling of SRAM dynamic stability.," in *IEEE/ACM Int. Conf. on Computer-Aided Design*, Nov. 2006.

[76] H. Pilo et al, "An SRAM design in 65nm and 45nm technology nodes featuring read and write-assist circuits to expand operating voltage," in *Symp. on VLSI Circuits*, June 2006.

[77] Y. Wang, E. Karl, M. Meterelliyoz, F. Hamzaoglu, Y.-G. Ng, S. Ghosh,L. Wei, U. Bhattacharya, and K. Zhang, "Dynamic behavior of SRAM data retention and a novel transient voltage collapse technique for 0.6 V 32 nm LP SRAM," in *IEEE IEDM 2011*, 2011, pp. 32.1.1 - 32.1.4.

[78] S. Mukhopadhyay et al., "Capacitive coupling based transient negative bit-line voltage (Tran-NBL) scheme for improving write-ability of," in *ISCAS*, 2008, pp. 384-387.

[79] C. Calligaro, V. Liberali and A. Stabile., "A radiation hardened 512 kbit SRAM in 180 nm CMOS technology," in *Electronics, Circuits, and Systems*, Dec. 2009, pp. 655-658.

[80] and Dimitri A. Antoniadis Mark S. Lundstrom, "Compact Models and the Physics of Nanoscale FETs," *IEEE Trans. Elec. Dev.*, vol. 61, 2014.

[81] Synopsys, "Three-dimensional Simulation of 14/16 nm FinFETs With Round Fin Corners and Tapered Fin Shape," *White paper*, 2013.

Appendix

Modeling TG FinFET in the Ballistic Regime

In this section we extend the existing model of the ballistic double-gate structure (D. Jiménez, J. J. Sáenz, B. Iñíquez, J. Suñé, L. F. Marsal et al., 2003), to the tri-gate structure. The most challenging issue about modeling of the tri-gate FinFET is its complex electrostatics, since it requires analysis in the 3D due to the lack of symmetry, which hinders a lot of progress for more investigation of its performance through compact modeling. Various works (H. Abd El Hamid, J. Guitart, V. Kilchytska, D. Flandre, and B. Iniguez, Sep. 2007), (G. Pei, J. Kedzierski, P. Oldiges, M. Ieong, and E. C.-C. Kan, Aug. 2002)was developed to model such complex geometries yet, most of them were restricted for some ranges of operation and under certain conditions. Here, we use a recently developed universal core charge model (J.P. Duarte, et al., , 2013) that calculates the charge for different multiple-gate structures including the DG and TG that is valid over the whole range of operation. First, we examine this universal charge model for the existing ballistic DG model. Hence, we use its TG charge formulation to develop new ballistic TG model. The well-known approach to model a current in nano-scale device, so far, that has been widely used for all ballistic models previously developed is by applying the Landauer formalism (R. Landauer, 1989) at the top of the barrier (the virtual source point) based on its transmission theory. In a same manner, as the DD model is characterized by some fundamental parameters like the mobility and the diffusion coefficient Landauer model is also characterized, however, by different parameters which are the transmission probability at energy E: T(E); for fully ballistic transport T(E)=1, and the number of current-carrying channels at energy E: M(E). The conceptual algorithm to model any ballistic device can be depicted in a flow chart as shown in Fig.1.

The Landauer formula can be developed to express the current this way

$$I = \frac{q}{\pi\hbar}\sum_{v}\sum_{n} g_v \left\{ \int_{E_x} [f_S(E) - f_D(E)]T(E)dE_x \right\} \quad (1)$$

where the inner sum with sub-index n indicates the discrete energy subbands, and the outer sum with subindex v represent the valley, and the quantity $q/\pi\hbar$ is the current carried per occupied subband per unit energy. Generally, we have confinement in two directions, and the electrons can move in one direction, the transport direction, hence, the energy E can be expressed in terms of two components, the one in the transport direction E_x which is continuous, assuming parabolic dispersion relation, and the component in the confinement zy-plane.

As it can be inferred from equation (1) that the calculation of the current encounters evaluation of Fermi-Dirac integrals

$$I = \frac{q}{\pi\hbar}\sum_{v}\sum_{n} g_v\{\Im_0(\eta_{F1}) - \Im_0(\eta_{F2})\} \qquad (2)$$

$$\eta_{F1} = \frac{E_{FS} - E_n^v(x_{max})}{KT}, \eta_{F2} = \frac{E_{FD} - E_n^v(x_{max})}{KT} \qquad (3)$$

where $E_n^v(x_{max})$ are the discrete energy subbands with respect to the bottom of the conduction band at the top of the barrier, taking the first subband $E_1(x_{max})$ as a reference, when evaluating the location of the Fermi-level, η_{F1}, η_{F2} can be rewritten in this way, (D. Jiménez, J. J. Sáenz, B. Iñíquez, J. Suñé, L. F. Marsal et al., 2003)

$$\eta_{F1} = \left(\frac{\left(E_{FS} - E_1(x_{max})\right) - \left(E_n^v(x_{max}) - E_1(x_{max})\right)}{KT}\right) \qquad (4)$$

Substituting into the current equation, we get

$$I = \frac{qKT}{\pi\hbar}\sum_{v}\sum_{n} g_v\left\{\Im_0\left(\frac{\left(E_{FS} - E_1(x_{max})\right) - \left(E_n^v(x_{max}) - E_1(x_{max})\right)}{KT}\right) - \Im_0\frac{\left(E_{FD} - E_1(x_{max})\right) - \left(E_n^v(x_{max}) - E_1(x_{max})\right)}{KT}\right\} \qquad (5)$$

From equation (5), to evaluate the current we need to solve for two quantities, one is bias independent $\left(E_n^v(x_{max}) - E_1(x_{max})\right)$, which are the separation between the upper subbands and the first subband, and the second is bias dependent $(E_{FS} - E_1(x_{max}))$ which involves solving for the electrostatics.

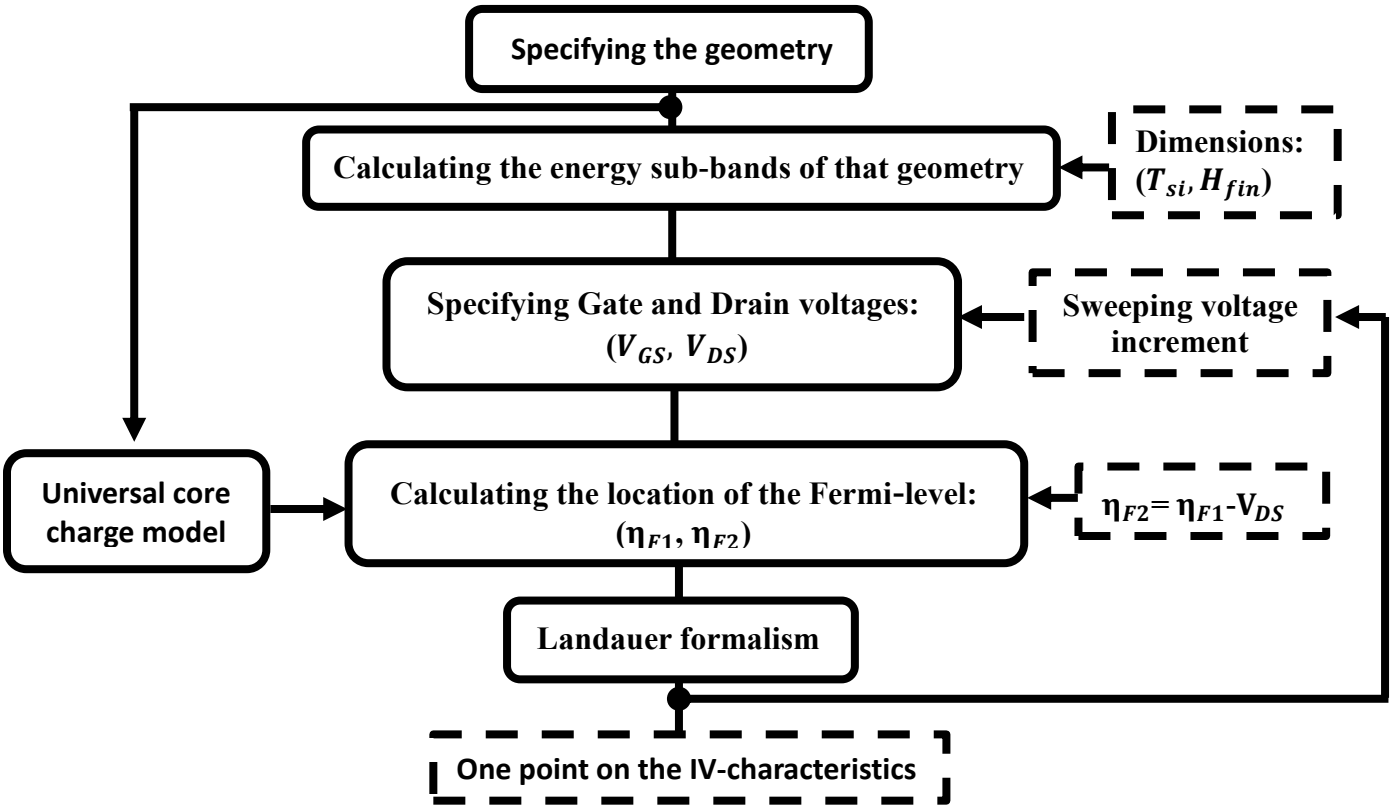


Figure A.1: Universal Flow chart to model Ballistic Transport for Multi-gate Structures

Starting with the bias independent term, calculating the energy subbands due the quantum confinement depends on the geometry of structure under study. For the TG structure, the electrons can be assumed to be confined in a two dimensional infinite square well hence, the separation between the bottom of the conduction band and the bottom of the sub-bands can be given by, (J. H. Davies, 1998)

$$\Delta E^{v}_{n_y,n_z} = \frac{\hbar^2}{2}\left\{\frac{(\pi n_y)^2}{m^v_y {T_{si}}^2} + \frac{(\pi n_z)^2}{m^v_z {H_{fin}}^2}\right\} \qquad (6)$$

where $(m^1_y, m^2_y, m^3_y) = (m_T, m_L, m_T), (m^1_z, m^2_z, m^3_z) = (m_L, m_T, m_T)$; the effective masses. For the DG structure, at the limit of very large height (H_{fin}), the confinement in the two directions reduces to a confinement in only one direction, and the two dimensional infinite square well reduces a 1D quantum well.

Moving to the electrostatics part to calculate the bias dependent term and starting with the DG structure, the charge distribution in (D. Jiménez, J. J. Sáenz, B. Iñíquez, J. Suñé, L. F. Marsal et al., 2003) was computed based on the boundary condition at the silicon oxide interface, the mobile charge can be written as:

$$Q = 2C_{ox}(V_{GS} - \phi_{ms} - \psi_s) \qquad (7)$$

$$\phi_{ms} = \phi_m - (\chi_{Si} - E_g/2q) \qquad (8)$$

where $\phi_{ms}$is the work-function difference, and ψ_s, the surface potential, is function of V_{GS} and can be solved iteratively as reported in (Yuan Taur , 2000), (Yuan Taur , 2001).

Here, instead, we will solve for Q directly using the universal core charge model (J.P. Duarte, et al., , 2013). The universal charge equation is

$$V_{GS} - V_{FB} + \frac{Q_{d,n}}{C_{g,n}} - V = -\frac{Q_{e,n}}{C_{g,n}} + v_T ln\left[\frac{-Q_{e,n}}{q\frac{n_i^2}{N_{Si}}A_{ch,n}} \frac{-(Q_{e,n}-Q_{d,n})/v_T C_{ch,n}}{1-exp\frac{Q_{e,n}+Q_{d,n}}{v_T C_{ch,n}}}\right] \qquad (9)$$

where $Q_{d,n}$ is the depletion charge per unit length, $Q_{e,n}$ is the mobile electron charge per unit length, $C_{g,n}$ is the gate oxide capacitance per unit length, $C_{ch,n}$is the channel capacitance per unit length, $A_{ch,n}$is the area of the channel.The subindex "n" denotes the used device architecture.

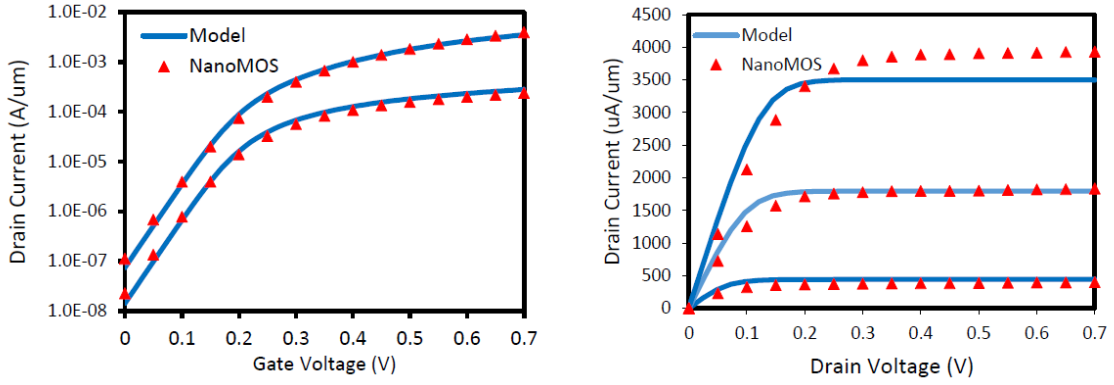


Figure A.2: I-V characteristics of DG structure of 20nm channel length, and 3nm silicon film thickness: (a) Transfer characteristics at VDS=0.05V, and 0.7V; (b) Output characteristics at VGS=0.3V, 0.5V, and 0.7V.

For the double gate structure (n=DG), this set of parameters is used

$$Q_{d,DG} = -qN_{Si}W_{Si}T_{Si} \quad (10)$$

$$C_{g,DG} = 2W_{Si}\frac{\epsilon_{ox}}{t_{ox}} \quad (11)$$

$$C_{ch,DG} = 4W_{Si}\frac{\epsilon_{Si}}{T_{Si}} \quad (12)$$

$$A_{ch,DG} = W_{Si}T_{Si} \quad (13)$$

hence we can solve for the mobile charge density Q_e. Also the charge can be expressed in terms the same unknowns as the current equation (D. Jiménez, J. J. Sáenz, B. Iñíquez, J. Suñé, L. F. Marsal et al., 2003)

$$Q_e = \frac{q\sqrt{2KT}}{2W\pi\hbar}\sum_v \sum_n g_v \sqrt{m_d^v}\; x\left\{\Im_{-1/2}\left(\frac{\left(E_{FS}-E_1(x_{max})\right)-\left(E_n^v(x_{max})-E_1(x_{max})\right)}{KT}\right)+ \Im_{-1/2}\left(\frac{\left(E_{FD}-E_1(x_{max})\right)-\left(E_n^v(x_{max})-E_1(x_{max})\right)}{KT}\right)\right\} \quad (14)$$

Substituting with the obtained charge from the universal core model into equation (5), we can solve for bias dependent term $(E_{FS} - E_1(x_{max}))$ at a given bias, knowing that

$$E_{FD} = E_{FS} - qV_{DS} \quad (15)$$

Now we are ready to substitute in equation (1) and get the IV-characteristics for the DG structure. Comparing the results with corresponding ballistic simulations from NanoMOS self-consistent 2-D simulator (Zhibin Ren; Sebastien Goasguen; Akira Matsudaira; Shaikh S. Ahmed; Kurtis Cantley; Yang Liu; Mark Lundstrom; Xufeng Wang (2013), "NanoMOS," https://nanohub.org/resources/nanomos. (DOI: 10.4231/D3J96090H).), choosing a test case structure of 20 nm channel length, 3 nm silicon film thickness, gate oxide of 1.5 nm thickness, and gate work function of 4.25 eV, for high performance (HP) device, we notice as shown in Fig.2 a good agreement between the two models. The next step is to extend the model for the TG structure.

The following modifications need to be incorporated:

(i) The confinement is in two directions, equation (1).

(ii) Tri-Gate charge formulation, given by substituting with the suitable set of parameters (J.P. Duarte, et al., , 2013)

$$Q_{d,TG} = -qN_{Si}H_{Si}W_{Si} \quad (16)$$

$$C_{g,TG} = \frac{3.02 \times 3\varepsilon_{ox}/2}{\ln(1 + 3t_{ox}/2H_{Si})} - \frac{5\varepsilon_{ox}/4}{\ln(1 + 5t_{ox}/4H_{Si})} + \frac{5\varepsilon_{ox}/4}{\ln(1 + 5t_{ox}/4W_{Si})} \quad (17)$$

$$C_{ch,TG} = W_{Si}\frac{\varepsilon_{Si}}{H_{Si}} + 4H_{Si}\frac{\varepsilon_{Si}}{W_{Si}} \quad (18)$$

$$A_{ch,TG} = H_{Si}W_{Si} \quad (19)$$

into equation (). Similarly, we choose test cases for the TG structure, channel length of 20 nm, silicon thickness of 3 nm to curb short channel effects, and various fin heights of 27 nm, 8 nm, and 4 nm, gate oxide thickness of 1.5 nm, and gate work function of 4.25 eV. The corresponding predicted ballistic transfer characteristics of the TG structures are depicted in Fig.3, and the impact of the fin height on the characteristics is well shown as increase in the threshold voltage due to the quantum confinement with the reduction of the fin height, since the bottom of the subbands in each silicon valley increases as expected from the quantum theory (Jean-Pierre Colinge, Fellow, IEEE, John C. Alderman, Weize Xiong, Member, IEEE, and C. Rinn Cleavelin, 2006). Fig.4 shows the increase of the lowest energy subband in each valley, with respect to the bottom of the conduction band, along with the threshold voltage increase, where the threshold voltage is defined here as the gate voltage that is required to bring the lowest subband in line with the equilibrium Fermi-level at the source end; i.e., to make the bias dependent term in equation (1), $(E_{FS} - E_1(x_{max}))$, equals to zero.

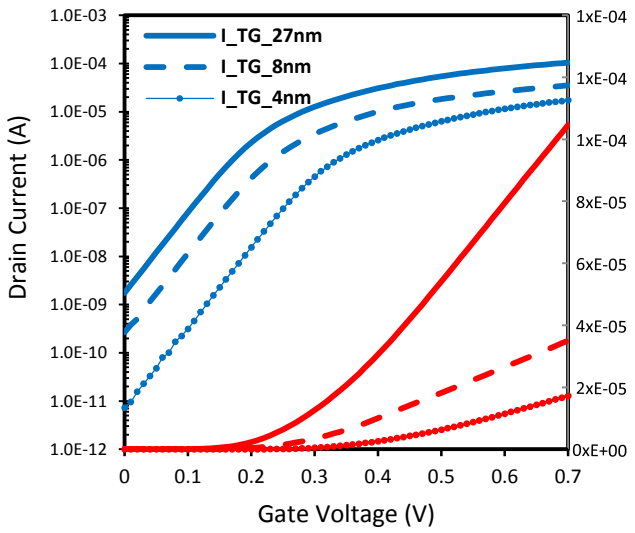


Figure A.3: Predicted ballistic transfer characteristics of TG-FinFET structures at drain bias of 0.8 V for fin heights of 27 nm, 8 nm, and 4 nm.

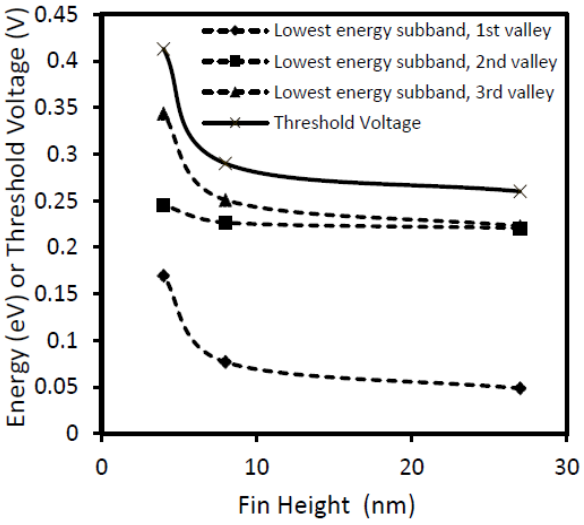


Figure A.4: The variations of the lowest energy subband as a function of the fin height

I. References

7. D. Jiménez, J. J. Sáenz, B. Iñíquez, J. Suñé, L. F. Marsal et al. (2003). Unified compact model for the ballistic quantum wire and quantum well metal-oxide-semiconductor field-effect-transistor. *J. Appl. Phys. 94, 1061*.

8. G. Pei, J. Kedzierski, P. Oldiges, M. Ieong, and E. C.-C. Kan. (Aug. 2002). FinFET design considerations based on 3-D simulation and analytical modeling . *IEEE Trans. Electron Devices, vol. 49, no. 8, pp. 1411–1419*.

9. H. Abd El Hamid, J. Guitart, V. Kilchytska, D. Flandre, and B. Iniguez. (Sep. 2007). A 3-D analytical physically based model for the subthreshold swing in undoped trigate FinFETs. *IEEE Trans. Electron Devices, vol. 54, no. 9, pp. 2487–2496*.

10. J. H. Davies. (1998). The Physics of Low-Dimensional Semiconductors. *Cambridge: U.K.*

11. J.P. Duarte, et al., . (2013). A Universal Core Model for Multiple-Gate Field-Effect Transistors. Part I: Charge Model. *IEEE Trans. Electron Devices, v. 60, p. 848*.

12. Jean-Pierre Colinge, Fellow, IEEE, John C. Alderman, Weize Xiong, Member, IEEE, and C. Rinn Cleavelin. (2006, MAY). Quantum–Mechanical Effects in Trigate SOI MOSFETs. *IEEE TRANSACTIONS ON ELECTRON DEVICES, VOL. 53, NO. 5*.

13. R. Landauer. (1989). Conductance determined by transmission: probes and quantised constriction resistance. *J. Phys. Condens. Matter, 1, pp. 8099-8109*.

14. Yuan Taur . (2000). An analytical solution to a double-gate MOSFET with undoped body. *Electron Device Letters, IEEE, 21*(5), 245- 247 .

15. Yuan Taur . (2001). Analytic solutions of charge and capacitance in symmetric and asymmetric double-gate MOSFETs. *Electron Devices, IEEE Transactions on , 48*(12), 2861- 2869 .

16. Zhibin Ren; Sebastien Goasguen; Akira Matsudaira; Shaikh S. Ahmed; Kurtis Cantley; Yang Liu; Mark Lundstrom; Xufeng Wang (2013), "NanoMOS," https://nanohub.org/resources/nanomos. (DOI: 10.4231/D3J96090H). (n.d.).